



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Acting in a populated environment:
An ecological realist enquiry into speaking
and collaborating**

Ed Baggs



Doctor of Philosophy

Institute for Language, Cognition and Computation

School of Informatics

University of Edinburgh

2015

Abstract

The thesis seeks to develop an account of collaborative activities within the framework of ecological realism—an approach to psychology developed by James J. Gibson in the course of work on visual perception. Two main questions are addressed; one ontological, and one methodological. The ontological question is: given that collaborative activities take place within an environment, what kinds of structure must this environment contain? The response emphasizes the importance of *relations* which exist between entities, and which connect a given perceiver-actor with the other objects and individuals in its surroundings, and with the relations between those entities. It is held that activities take place within a field of relations. This description draws on the radical empiricist doctrine that relations are real, are external, and are directly perceivable. The present proposal insists that, in addition to being directly *perceivable*, relations can also be directly *acted upon*: throwing a ball for a dog is acting on a relation between dog and ball in space. The relational field account of collaboration naturally extends to an account of speaking: people, through their history of acting in an environment populated by other speakers, come to stand in a set of relations with objects and events around them, and these relations can be directly acted upon by others through the use of verbal actions. Verbal actions serve to direct the attention of others to relevant aspects of the environment, and this allows us as speakers to coordinate and manage one another's activity.

The methodological question is this: granting that the environment may be structured as a field of relations, how are we to conduct our empirical investigations, such that we can ask precise questions which lead to useful insights about how a given collaborative activity is carried out in practice? The central issue here concerns the concept of the *task*. Psychologists are in the habit of using this term quite loosely, to denote the actions of an individual or a group, in a laboratory or outside. This creates confusion in discussions of collaborative phenomena: who is the agent of a 'collaborative task'? The definition offered here states that a task is a researcher-defined unit of study that corresponds to a change in the structure of the environment that has a characteristic pattern and that is meaningful from the first-person perspective of a particular actor. On this definition, the task is a tool that allows ecological psychologists to carve up the problem space into specific, tractable questions; the task is the equivalent

of the cognitivist's mental module. Task-oriented psychology encourages us to ask the question: which specific resources is the individual making use of in controlling this particular activity?

The methodology is developed through an examination of the alarm calling behaviour of vervet monkeys, which is explained in terms of actions on the relational field, and through an analysis of corpus data from a laboratory-based collaborative assembly game. The relational field model promises to provide a way of studying social and collaborative activities on ecological realist principles. The concluding chapter identifies two particular areas in which the model might fruitfully be developed: in the study of learning, and in the theory of designing objects and spaces for interaction.

Lay summary

The thesis attempts to develop an account of how some animals, including humans, are able to engage in activities together. I approach this question not by asking about features of the actors' brains, but by noting that the environment surrounding a given individual is populated with other actors, and by seeking relevant features of that environment that may support action. I outline, first, a scheme for describing the environment surrounding a given actor, and, second, a methodological programme for studying the behaviour of actors in populated environments. In the descriptive part, the environment is conceived as a richly structured network of relations in which the intentions of one's fellow actors can be perceived over time in the unfolding of the activities in which those actors are engaged. I use this scheme to outline an account of language learning in which young children learn to speak by first attending to the consequences of the sounds that they themselves make, and specifically to the effects of those sounds on the behaviour of other actors in the child's environment. I relate this account to existing research on spoken dialogue.

The methodological part of the thesis proposes that in order to study behaviour in complex multi-actor settings it is first necessary to clarify the notion of the 'task'. I propose that a task is a unit that is defined by a researcher for the purpose of being able to study it. A task involves a reproducible structured pattern that appears in the environment of an actor and that is meaningful to that actor. On this definition, the task is a tool that allows psychologists to carve up the overall problem space into smaller, more manageable units. It allows us to ask not, 'What explains the actor's behaviour in general?' but 'What precise structure is the actor using to control this particular action in this specific context?' I identify two particular areas where the model developed might be fruitfully applied. First, in education, where learning can be understood in terms of a learner's developing ability to make use of structure inherent to a particular, well-defined task. And second, in design, where user experience can be understood in terms of the task structure that arises whenever the actor tries to achieve some particular outcome (e.g., the experience of vehicles approaching from a distance that arises when a pedestrian attempts to cross a road).

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(Ed Baggs)

Table of Contents

Preface	x
Acknowledgements	xiii
I Ontology: how is a populated environment structured?	1
1 A brief introduction to ecological realism	2
1.1 The ecological approach to perceiving and acting	3
1.1.1 The animal-specific environment v. the physical world	4
1.1.2 The concept of the invariant	5
1.1.3 Ecological optics v. physical optics	8
1.1.4 The concept of information	10
1.1.5 The concept of affordances	11
1.1.6 The role of learning: the doctrine of the new ball	13
1.2 The radical empiricist ontology	15
1.2.1 The reality of relations	17
1.2.2 The animal in a field of relations	18
1.2.3 The animal as a builder of its environment	19
1.3 Summary	21
2 Ecological realism socialized: Acting in a populated environment	22
2.1 Against ‘joint action’	23
2.2 Acting in a populated environment	26
2.2.1 Soccer	26
2.2.2 Chimpanzee pack hunting	28

2.3	Growing into one's environment: lessons from Holt and Vygotsky . . .	30
2.3.1	Holt and the recession of the stimulus	30
2.3.2	Vygotsky and the zone of proximal development	32
2.4	Summary	34
3	An ecological realist guide to speaking	35
3.1	Does 'language' exist?	37
3.2	Reconciling speaking and ecological realism	39
3.3	Speaking as acting on the relational field	45
3.3.1	Action controlled with reference to relations	48
3.3.2	Hearing: acting or being acted upon?	49
3.4	An outline of a theory of language learning	51
3.5	Putting the relational field model of language to work	57
3.6	Postscript: Columbus crosses the ocean	57
II	Methodology: studying activity in a populated environ-	60
ment		
4	The task-oriented approach in psychology	61
4.1	The task as an epistemological device	63
4.2	The acting animal and the task-seeking psychologist	65
4.3	The task as a reconfiguration of resources	66
4.4	Acting with other actors	68
4.5	Rationale for the task-oriented approach	70
4.6	Final definition: what is a task?	72
5	A relational field-based account of referential communication games	75
5.1	The structure of referential communication games	78
5.1.1	The extended conduit model	79
5.1.2	The field model	85
5.1.3	A note on roles and repertoires	90
5.2	Accommodating established findings	91
5.2.1	The reduction of referring expressions	92

5.2.2	The difficulty of monologue	95
5.2.3	‘Audience design’	96
5.2.4	The overhearer’s disadvantage	98
5.2.5	Clarification requests	100
5.3	Summary	102
6	The ecosystem of vervet monkey alarm calls	104
6.1	Some facts about vervet alarm calls	105
6.2	Alarm calls as actions in the relational field	107
6.3	Monkey alarm calls and human speaking	114
6.4	Summary	117
7	Action perturbation in a collaborative assembly game	119
7.1	Learning and perturbation	124
7.2	Perturbation in the JCT	126
7.3	What is the task in the JCT?	130
7.4	Perseverative wiggling	134
7.5	Performing by looking	137
7.6	Some open questions	140
7.7	Summary	143
III	Closing remarks	144
8	Lessons learned and prospects for future research	145
8.1	Things I have learned about the psychology of speaking and collaborating	146
8.2	Prospects for future research	154
8.2.1	Learning	155
8.2.2	Design	159
8.3	Exploring the field	164
Appendix A	A taxonomy of approaches to the analysis of collaborative activities	165
A.1	Overview	165

A.2	Internalist approaches	167
A.2.1	The group mind	167
A.2.2	Individualism	168
A.3	Distibuted cognition	171
A.4	Externalist approaches	172
A.4.1	Interpersonal synergies	172
A.4.2	Radical empiricism	173
Appendix B Glossary		175
Bibliography		190

*'There's nothing I admire more in
philosophy than a well-bitten bullet.'*

Jerry Fodor (2000)

Preface

Included in a posthumously published collection of the papers of James J. Gibson is a two-page memo outlining some of his thoughts on language (Gibson, 1982, pp. 411–412). In it, he asserts that language is a tool for indirect perception—a means of seeing at second-hand. I believe this view of language is ultimately incompatible with Gibson’s ecological realist programme as a whole, for reasons I will discuss in chapter 3. The argument will be that this view of language inherits from earlier philosophers an untenable dualism between language and the physical world. I mention the memo here because of its title: ‘Note on perceiving in a populated environment’. This, it seems to me, makes explicit a point that is of fundamental importance. To study how an animal collaborates with its fellows is to study how that animal behaves in an *environment* that is *populated*. This point is, of course, absolutely obvious. Yet it is one that is insufficiently appreciated and frequently forgotten by philosophers and psychologists attempting to explain collaborative activity. If I were to summarize the present thesis in one sentence, it might be this: *That the environment is populated is a fact that should not be forgotten*. The thesis aims to identify some principles for studying the collaborating animal in a manner that fully recognizes the fact that the animal occupies a place in an animate world. I have borrowed Gibson’s phrase for the title.

Naturally, there already exist several competing approaches to the study of the collaborating animal. It is not my aim to prove that these existing approaches are wrong, nor that the theoretical approach adopted here is the *only possible* correct one. To do so would be to construct what Chemero (2009) calls a ‘Hegelian argument’.¹

¹This phrase refers to an argument made by Hegel in relation to the possibility that astronomers might discover a new planet between Mars and Jupiter; he argued that such a discovery was impossible. According to Chemero (2009, p. 5), Hegel’s argument was not made on the basis of any empirical data drawn from observation of the heavens, but simply on the a priori belief that the order of the planets must necessarily conform to a certain number series described in Plato’s *Timaeus*. A ‘Hegelian argument’ is

I have, however, found it necessary to spend some time, particularly in chapters 1, 4, and 5, discussing the differences between the approach being outlined and some competing approaches. I hope that this emphasis on differences between approaches, where it exists, serves to make clear the metatheoretical assumptions that are being made.

The positive account that is set out in the following chapters is constructed within the framework of ecological realism. This is a coherent and well-established programme, with a well-defined set of basic principles, which differ in important ways from the principles adopted by other approaches. One such difference concerns the question of other minds. Most attempts to construct a psychology of collaborative activity run up against the ‘problem of other minds’: if minds are private entities which contain images or representations, how are we able to gain knowledge of other individuals’ minds, and how are we able to coordinate activities between ourselves? The ecological approach, though, rejects the notion of mental content: mental content is not among its basic principles. Within the ecological system, the problem of other minds simply does not arise. Attending to basic differences between competing approaches, then, is necessary if we wish to avoid blind alleys. The only way to discern which of the traditional problems are relevant for the ecological realist is to be clear about what our starting assumptions are.

A note on the structure of the text. The bulk of what follows is divided into two parts. The first part is intended to deal with some basic ontological issues: what kinds of causal machinery must exist in the world if an ecological account of collaboration is to be viable? The second part aims to develop an empirical methodology: how can we use the causal machinery we have identified to ask questions about how actors go about achieving specific tasks? The reader may note that this division does not always hold—that the first part contains some discussion of methodological issues, and that the second introduces a good deal of ontological detail that was not mentioned in the first part. It is hoped that such departures from the grand structure will not detract unduly from the reader’s enjoyment of the text. Part three consists of a closing chapter which

an argument ‘based on little or no empirical evidence’ that some scientific research programme is futile (p. 7). Chemero sets out a persuasive case that such arguments are unproductive and that a scientific research programme should be judged not on how sensible it seems at the outset but on its success over the long term.

does not attempt to summarize parts one and two but instead picks up on some under-examined themes about the practice of research, and then looks forward to potential ways in which the ecological account of collaboration might be further developed and applied.

Acknowledgements

The work carried out in this thesis was supervised by Holly Branigan and Jon Oberlander. I could not have asked for a better pair of supervisors. Their openness to dissent and intolerance of unsound logic have shaped the ideas contained herein into something more complete and interesting than I had thought possible.

The work has also benefitted from many thoughtful discussions at the Joint Action Reading Group. I thank everyone who has attended these meetings, in particular: Chiara Gambi, Olle Blomberg, Alessia Tosi, Alex Papadopoulos, Jarek Lelonkiewicz, Greg Mills, and Sophia Lycouris. And thanks also to Chiara and Alex for reading parts of the manuscript.

I also thank Rachael Bailes, Justin Quillinan, and Bill Thompson. And Robin Hill, who helped with the material in chapter 7, on the JCT. The work was influenced at various stages by comments from Martin Pickering, John Lee, and Alex Lascarides, and by discussions with Kerry Marsh, Danielle Matthews, and Pat Healey.

I must also acknowledge the influence of two blogs, both of which have stimulated the present work in important ways, although this is hardly apparent from their authors' presence in the bibliography. Eric Charles's blog 'Fixing psychology' covers many deep issues to do with the foundations of ecological psychology; it is from here that I learned about E. B. Holt and the 'new realists'. And Andrew Wilson and Sabrina Golonka's blog 'Notes from two scientific psychologists' does an excellent job of defending ecological psychology from its critics and of demonstrating the empirical effectiveness of the approach. I thank Eric for his helpful comments on a version of chapter 3 that has now been published. I thank Andrew and Sabrina for kindly agreeing to discuss with me some of the ideas that follow.

Part I

**Ontology: how is a populated
environment structured?**

Chapter 1

A brief introduction to ecological realism

Following [Michaels and Carello \(1981\)](#), I will use the term *ecological realism* as a label for the ontological framework within which the present thesis is written. The aim of this first part of the thesis is to outline the existing ontology of that framework and to describe how that ontology can be applied to collaborative activities. This first chapter briefly identifies the central tenets of the ecological approach. I do not attempt a complete defence of the approach (a task that has been ably attempted elsewhere).¹ Nor do I attempt an exhaustive overview of everything that ecological psychologists have to say. Rather, the aim of the present chapter is to provide a set of orienting tools that will be necessary in what will follow. Specifically, I aim to identify the most important characteristics of ecological realism that distinguish it from the mainstream, i.e. cognitivist, understanding of psychology, and to draw particular attention to overlapping terminology which is apt to obscure those differences and cause confusion. Note that I have also provided a glossary in [Appendix B](#). The first part of this chapter will introduce the ecological approach to perceiving and acting. The second part will introduce the idea of an ecosystem as a network of relations within which the animal is situated; this second part will identify some tools that will be called upon in later chapters, as

¹For a more comprehensive and highly readable introduction to the philosophy of ecological realism, I refer the reader to [Michaels and Carello's *Direct Perception* \(1981\)](#). The present chapter is also informed by more recent books ([Shapiro, 2011](#); [Chemero, 2009](#); [Barrett, 2011](#)), and by Gibson's own writing ([Gibson, 1966, 1979](#)).

we attempt to develop a way to incorporate collaborative activities into the ecological realist programme.

1.1 The ecological approach to perceiving and acting

James J. Gibson developed his ecological approach to psychology in the course of his work on perception—principally visual perception (Gibson, 1966, 1979). The study of perception remains the area in which the ecological approach is most thoroughly developed. Gibson conceived of perceiving as an active process that extends through time as an animal constantly explores its environment. The environment is said to be richly structured, and to be perceived directly, that is, without the mediation of mental pictures or symbols, and without the need for the animal to carry out any inferential processing of the ‘input’. Perceiving is further held to be indivisible from acting:² the targets of perceiving are never pure forms, but are *affordances*, or opportunities for the animal to act that arise from correspondences between the animal’s body and appropriately-sized structures in the animal’s environment.

It is important to grasp that Gibson’s approach to psychology is quite different from the standard cognitivist approach. Crudely put, the standard approach starts with the assumption that the world exists as an external reality, and asks what it is about the *internal* structure of the animal that allows the animal to behave adaptively with reference to that reality.³ Gibson’s approach starts instead from the position that the animal and its environment constitute, from the beginning, a single, coherent system. Gibson focuses his attention on the structure that stands *between* the animal and its surroundings; he asks: what is it about this *relational structure* that enables the animal to successfully coordinate its behaviour?

²For a related argument, see Noë’s (2004) enactive account of perception, according to which perception is itself conceived as a kind of action.

³For a sympathetic outline of cognitivism, see Haugeland (1978): ‘Cognitivism in psychology and philosophy is roughly the position that intelligent behavior can (only) be explained by appeal to internal “cognitive processes,” that is, rational thought in a very broad sense.’

1.1.1 The animal-specific environment v. the physical world

At the very start of his final book, Gibson introduces his concept of the environment by contrasting it with the standard view of the physical world. The terms ‘environment’ and ‘physical world’, he claims, should not be treated as synonyms. Understanding this distinction is key to understanding the radical difference between Gibson’s psychology and the standard approach. (The distinction is evidently very important to Gibson: he makes it on page 2 of chapter 1, a chapter whose title is “The animal and the environment”).)

The environment, according to Gibson, is that which surrounds an individual animal (1979, p. 8):

[I]t is often neglected that the words animal and environment make an inseparable pair. Each term implies the other. No animal could exist without an environment surrounding it. Equally, although not so obvious, an environment implies an animal (or at least an organism) to be surrounded. This means that the surface of the earth, millions of years ago before life developed on it, was not an environment, properly speaking. The earth was a physical reality, a part of the universe, and the subject matter of geology. [...] We might agree to call it a world, but it was not an environment.

Gibson is here insisting that animals do not perceive the world of physics, we do not perceive the bare physical world. What we perceive is our surroundings: the structured collection of surfaces and objects that are meaningful by virtue of our capacity to act relative to them. The *environment* ultimately only exists at the approximate scale of our bodies. (This is true even if we have invented such instruments as microscopes and mass spectrometers that allow us to peer into the structure of the world at scales that we are not evolved to perceive directly.) When ecological psychologists talk about the ‘animal–environment system’, this is not some vague device for pointing out that our bodies are situated in space and that we perceive things in that space (a fact which no cognitivist would deny). It is, rather, an assertion that we never in fact perceive pure ‘space’. What we perceive is our surroundings; we perceive our own unique environment relative to our ability to act in it.

1.1.2 The concept of the invariant

To point out that the animal has its own unique set of surroundings is not sufficient, of course, to explain how it is that that animal is able to act relative to those surroundings. In order to move in that direction, we need take up the concept of the *invariant*. This is perhaps the most important concept within ecological psychology. It is this concept that allows ecological psychology to provide a viable account of perceiving as a process unmediated by mental images—as a process that is direct, active, and *for* the detection of affordances.

For Gibson, the great mistake of previous approaches to visual perception was the idea that perceiving is reconstructing a picture of the world from the distorted image that is projected onto the retina. It is easy to make this mistake. We are encouraged by the ubiquitous metaphor of the eye as a camera. In fact, the eye is not a camera, and the retina is not a photographic plate: the retina does not produce a very useful photograph and even if it did, there would normally never be anybody there to see it (the only way in which a retinal image *can* be perceived is by removing the eyeball from an animal, shining a light through it, and observing it from behind—a feat once demonstrated by René Descartes, who used the eye of a bull). Gibson concluded that the retinal image, at least as a causal component in perceiving, is a fiction. Perceiving cannot be about reconstructing an imperfect image, but must be about something else. He recognized that an alternative account was needed, and set out to provide that account.

The account he devised says that light—the kind of light that is present in an animal's evolutionary environment—contains structure which is specific to the animal's surroundings, structure which the animal can learn to respond to in increasingly sophisticated ways. Contrary to the assumption that stimulus-data is impoverished—the assumption that motivates the traditional view of perceiving as reconstructing (e.g., [Fodor, 1983](#))—this alternative account says that the stimulus is in fact richly structured, sufficiently so that the stimulus itself can be used to guide action.

Further, this structure can be divided into two parts: structure that changes under transformation as the animal moves about, and other structure that does not change. For example, when a flat rectangular surface is viewed from an angle, it projects a different shape to the viewer's point of observation than it does when viewed from

directly above. The act of moving towards the rectangle causes this projected shape to transform. At the same time, there are properties of the projected shape that do not change as one walks towards the rectangle. The internal angles of the shape remain constant, as does the number of sides, the positions of the sides ABCD relative to one another, the surface texture, the colour. Gibson's insight was to suggest that these unchanging aspects of the stimulus (the patterns in light) are the kinds of structure that are reliable enough to underpin the control of action. This unchanging structure is *invariant structure*: structure that specifies the layout of the animal's environment. Notice that in order to make use of this structure, the animal must already be able to move about. This meant that for Gibson the stationary image was in fact a special case of perception in motion, not the other way around. Gibson had turned the traditional approach on its head.

To see how difficult it is to grasp this inversion of thinking, it is interesting to note how researchers in the cognitivist tradition have attempted to respond to Gibson's proposals. We will take one example here. David Marr, in his influential book *Vision* (1982), recognized in Gibson's account an attempt at a computational-level description (in Marr's terms) of what the problem is that the perceiving animal must solve (p. 29):

'In perception, perhaps the nearest anyone came to the level of computational theory was Gibson (1966). However, although some aspects of his thinking were on the right lines, he did not understand properly what information processing was, which led him to seriously underestimate the complexity of the information-processing problems involved in vision and the consequent subtlety that is necessary in approaching them.'

According to Marr, Gibson's proposals suffered from two fatal shortcomings: 1) Gibson is too quick to reject information-processing as a mechanism: 'the detection of physical invariants, like [the detection of] image surfaces, is exactly and precisely an information-processing problem, in modern terminology', and 2) Gibson underestimates how difficult this problem is: 'Detecting physical invariants is just as difficult as Gibson feared, but nevertheless we can do it. And the only way to understand how is to treat it as an information-processing problem.'

What appears to be going on here is that Marr has misinterpreted the meaning of 'invariant', as Gibson uses the term. Marr takes it merely as a label for the fact that *objects* remain the same shape and size over time, i.e. as a synonym for 'constancy'.

In other words, Marr is interpreting the word as referring to a property of the external environment, while Gibson intends it to refer to a property of the *stimulus*, i.e. the structured light that is intercepted by an exploring animal (Gibson, 1960). For Gibson, the animal perceives the structure of the environment by attending to invariant structure in light. This structure specifies the layout of the environment as a consequence of the lawful way in which light is reflected from surfaces: light is structured by the environment. Consequently there is no information-processing problem: the structure is already there in the light. Or put another way, the structure in light is already ‘processed’ at the point at which it is intercepted by the animal, by the fact that it is reflected in a lawful way from the environment’s surfaces. The task for the animal is to learn how to make use of this structure in order to control its actions.

What ecological psychologists mean when they refer to ‘an invariant’ is this kind of detectable property in the higher-order structure of light (or another medium): structure that guides action. A simple example is the invariant that enables an animal to move towards a stationary object. In this case, what the animal needs to do is to focus the intended destination in its centre of optical expansion as it moves (Gibson, 1958). As the animal moves towards the object, the object itself gradually expands to fill a larger portion of the animal’s visual field, while the surrounding surfaces appear to shear off to one side and eventually disappear from the field of vision altogether (at least for an animal with forward-facing eyes). Thus the position of the target object is constant during this particular act of locomotion, while structure that is not being used in the control of this action is allowed to vary: everything else blurs off to the side. Meanwhile, if the target object itself begins to blur to the side, this constitutes information for the animal that it has veered off course and needs to correct its direction of travel. This is what is signified by ‘an invariant’: a feature of the stimulus (light) that is revealed through motion and that can be reliably used to control a particular action.

Marr’s misinterpretation usefully illustrates two things. Firstly, it highlights that the term *invariant* is one that needs to be introduced with some care. And secondly, once the term has been correctly defined, it becomes clear that Marr and Gibson are simply pursuing different hypotheses about how vision works. What does it mean to say that apprehending an invariant in light is an information-processing problem? This

assertion makes sense only if we continue to insist upon the traditional conception of seeing as reconstructing a mental picture that is in turn perceived by a mind.⁴ But Gibson explicitly rejects this account of seeing. On Gibson's account, there is no need to reconstruct anything. Rather the structure is already there, or rather, it arises in the act of movement. The problem for the animal is to work out how to produce the relevant structure and to learn what that structure means.⁵

1.1.3 Ecological optics v. physical optics

Another way to try to make clear the deep differences between the ecological and the cognitivist approaches is in the distinction between classical or physical optics and Gibson's system of ecological optics (Gibson, 1961).

Physical optics is about measuring wavelengths and intensity of light, and about matching points in an image to corresponding points in a three-dimensional coordinate scheme: length, width, and depth. Here the measuring tools are themselves artifacts: rulers, protractors, light meters. These instruments are designed to give a reproducible measure of some property of the environment independent of how the environment is perceived subjectively. Additionally, the coordinate scheme itself is a formal tool for denoting specific points in physical space, again independent of subjective experience.

Ecological optics, by contrast, is about how the animal uses the structure available

⁴Marr (1982) does appear to allow that certain visual systems might function without the intervention of mental representations. He describes the landing system of the housefly thus (p.33): 'if the visual field "explodes" fast enough (because a surface looms nearby), the fly automatically "lands" towards its center.' This description is congruent with Gibson's (1958) description of the visual control of locomotion in humans. However, what Marr may be willing to allow for the fly, he is not willing to allow for humans. The fly's visual system is 'not very complicated' (p. 34). The human visual system, Marr asserts, is engaged in a different game, that of 'building a description of the shapes and positions of things from images'; the job of human vision is 'to derive a representation of shape' (p. 36).

⁵It is perhaps worth noting here that some ecological psychologists have made a distinction between two different types of invariant: 'structural invariants', and 'transformational invariants' (Pittenger and Shaw, 1975; Michaels and Carello, 1981). Both of these terms are supposed to denote a property of the stimulus (light). The former is supposed to denote structure in light that specifies inanimate objects (notice that this is still not the same thing as Marr's interpretation of an invariant as a property of the object itself), while the latter is supposed to denote structure in light that specifies events in the world. Personally, I find this distinction unhelpful: it seems to be one that is more likely to increase confusion rather than reduce it. After all, any act of perceiving is an event, whether or not the object being perceived is itself moving. This is not a distinction made by Gibson himself; in fact, this cuts to a deep difference in perspective between Gibson and some of his followers, a topic that is too arcane to get into here (Cutting, 1982).

in light to guide its own actions. For an animal, there is no formal requirement that an environment has to be perceived in exactly the same way from one point in time to another. The animal's perceptual systems—its measuring tools—are scaled not in the units of some abstract, formal system, but in terms of the actions the animal is currently capable of carrying out. The animal is not merely a perceiver, but a perceiver-actor: perceiving is inherently bound to the potential to act. Perceiving and acting are one and the same phenomenon.

What makes possible this alternative way of measuring light is the very fact that light is structured in a complex way. Rulers and light meters measure this structure in one manner: that is, they measure a particular aspect of the overall structure. But ecological optics says that the structure that animals use to guide their activity is not the same thing as the structure that happens to have been measured by the tools we have so far invented.

How is light structured? Ambient light is structured by the surfaces in an environment. Light enters the environment from a light source (the sun) and bounces around near-instantaneously from surface to surface such that, at the time scale of human experience, the environment is constantly filled with light from all directions. At any potential point of observation, the light that reaches that point is structured in a lawful way by the surrounding surfaces, and in a way that is unique to that point at that particular moment. The set of all such potential points of observation is referred to as the *ambient optic array*. (In the jargon of ecological psychology, the medium of perception is always an energy array. In the case of visual perception, the energy is light; for acoustic perception the energy is sound waves; and so on.)

This gives an indication of the richness of the structure available to perception. At any given point of observation, there is structure that specifies (structure in light that is specific to) the entire surroundings. Meanwhile, if an animal is unable to unambiguously perceive some part of its surroundings from a given stationary point, it need only move in some direction in order to generate for itself more information about the ambiguous structure. And since perceiving is not bound in time to a single moment, it is not a problem if the animal has to physically explore a part of its surroundings before it has satisfactorily perceived that structure.

1.1.4 The concept of information

Perhaps the term most apt to cause confusion within ecological psychology is the word ‘information’. One problem is that this is the same word as is used in the mainstream, information-processing account, where information is understood formally in terms of particular units or bits; that is, information is something that is held to exist independently of experience.⁶ By contrast, the ecological conception of information is again specific to a particular animal-environment system, and it does not make sense to talk about this kind of information except in the context of such a system: without animals, there is no information, there is only structure. Information is here defined as information-*for* an organism, as something incorporated into an ongoing activity that imposes constraints or direction on that activity (Michaels and Carello, 1981, p. 37).⁷ This conception of information is animal-relative, action-oriented, and content-free. Information arises in the course of the animal’s activity; it is *for* the detection of affordances, and not *about* the world outside in some generic sense. It makes no sense to say that light ‘carries’ information, although light is structured. This structure only becomes information when an animal detects it (van Dijk et al., 2015).

The habitual cognitivist will likely remain unconvinced by the assertion above that structure in light is simply available to a perceiving organism. In particular, it will likely be pointed out that to say that information is something that can be extracted from structure in light is to produce a fancy description while avoiding getting down to brass tacks. What is missing, it will be declared, is a plausible mechanism for how the nervous system turns structure in light into genuine constraints on action.

The challenge put forward by ecological psychology is to suggest that the nervous system by itself cannot be said to be *doing* anything. The nervous system does not, by itself, have agency. The nervous system is merely one component in a larger animal-environment system. To try to explain an animal’s behaviour purely from a description of its nervous system is like trying to explain the workings of a railway station from a description of the positions of the levers in the signal box.

⁶There are many ways of trying to define non-ecological information. I will not get into these here; I here confine myself to trying to summarize the nature of ecological information. See Adriaans (2013).

⁷The ecological realist proposal here anticipates a similar critique of the concept of information in biology: that it does not make sense to talk about information at all except relative to some process that makes use of the information (Oyama, 2000; Jablonka, 2002).

The position that ecological psychologists have often adopted on this problem is to say that in fact we cannot even begin to describe the activity within the nervous system until we have a satisfactory account of what the task is that the *animal-in-its-environment* is trying to accomplish (Wilson and Golonka, 2013). The animal-environment system is the unit of analysis; the working of any particular subcomponent, such as a nervous system, can only be assessed relative to the entire system. Another way in which this same point is put is that psychology, as a science, must operate at its own level of explanation, with its own set of primitives, and not with the same set of primitives as are used in physics, say (Michaels and Carello, 1981). Psychological primitives might include: the fact of subjective experience (Reed, 1996b), the fact that experience is inherently directed at external structure (and thus there is no problem of intentionality), and the fact of relations as perceivable entities (see below). This kind of thinking about psychology at its own level of explanation is what motivated the development of ecological optics in contrast to traditional physical optics (above).

At the psychological level of explanation, then, the nervous system is to be treated as a black box: a component somehow implicated in the conversion of structure into action. But if we step down to the level of the nervous system, the psychological explanation becomes a *description* of the task to be solved, and there are some preliminary points we can make here. Firstly, we can be confident that nervous system is *not* in the business of reconstructing a moving image of the world from impoverished sense data. Secondly, we can say, approximately, that the function that the nervous system *does* have has something to do with attuning the animal to invariants that are instrumental in constraining an ongoing action. This job description must also be informed by a set of properties at an even higher level of description, namely affordances.

1.1.5 The concept of affordances

The concept of the affordance is probably the concept most readily associated with ecological psychology by those outside the field, and in general it appears to be the case that those outside the field believe they understand what it means; cognitivists, for instance, are quite happy to import the concept into their own own explanatory toolbox

(e.g., [Kaschak and Glenberg, 2000](#)). In fact, however, by appropriating the term for use within their own scheme, cognitivists are radically distorting its meaning. The concept of affordances follows from the more basic tenets of ecological psychology: direct perception by an animal that must actively explore its richly structured environment.

In order to understand the concept it is necessary to first conceptualize the animal and its environment as constituting a system ([Stoffregen, 2003](#)). Just as perceiving and acting were characterized above as in fact two different ways of referring to a single phenomenon, namely perception-action, so the animal and its environment must be conceived as inherently inseparable: the one is strictly meaningless without the other, as Gibson insisted at the start of his book. An environment, recall, is defined here as the entirety of what surrounds an animal; it is not a synonym for the world.

Since animal and environment are conceived as a system, there must exist some mechanism by which the component parts fit together. Affordances provide such a mechanism. They are the relations that link an animal to particular assemblages of structure in its surroundings ([Chemero, 2003, 2009](#)). Specifically, they are relations that enable the animal to engage in perception-action. Affordances furnish opportunities for action and thus provide meaning to perception.⁸

Affordances, in the ecological scheme, pick out real correspondences between an animal's body and appropriately-scaled structures in the animal's environment. Some examples are self-enclosing surfaces whose diameter is no broader than the grip aperture of an animal's hand (a handle), or rigid surfaces that span a body of water and that are strong enough to support a specific animal (a bridge). These correspondences are specific to a particular animal and a particular piece of structure: what constitutes a handle for a chimpanzee does not constitute a handle for a tadpole or a parrot. These correspondences can be discovered, explored, and created. They physically provide and constrain the things an animal can do because they *are* the things an animal can

⁸Not everyone agrees that 'affordance' is the label of a relation; some writers, for instance, say that an affordance is a dispositional property of an external object, which is itself complemented by a corresponding property in the animal ([Turvey, 1992](#); [Fajen et al., 2009](#)). It has also recently been pointed out that the affordances-as-relations view is too broad, because it does not distinguish between action opportunities that merely exist in the animal's surroundings and those that enter into the animal's experience, or its directed activity, and which would seem to invite particular actions ([Withagen et al., 2012](#)). See [Käufer and Chemero \(2015\)](#) for further discussion. For present purposes, I simply adopt the affordances-as-relations position, and will use the term consistently throughout.

do.

In contrast, the cognitivist appropriation of the term appears to treat ‘affordance’ as a synonym for ‘category’: it treats perceiving affordances not as perceiving the actions that things afford, but rather as a special manner of perceiving objects, of perceiving them *as* things that afford particular actions. Cognitivists talk of constructing a mental model of an environment and exploring the affordances of that model (e.g., [Kaschak and Glenberg, 2000](#)). This, I suggest, is merely creating an unnecessary step, and moreover an entirely unhelpful step: it is changing affordances from tractable, directly measurable phenomena (actual correspondences between animals and physical structures) into opaque mental phenomena that can at best be inferred indirectly.

1.1.6 The role of learning: the doctrine of the new ball

All this talk of perceiving and acting as direct exploitation of a richly structured environment may appear to come up short when confronted with the fact that animals learn. If nothing ever really enters the animal in the form of mental content,⁹ how is it that animals learn new skills as they grow, and perform actions that appear to rely on remembered events?

To answer this it is necessary to recognize another deep difference in thinking between the ecological approach and the mentalist alternative. Here the question is: what is learning? The traditional answer is framed in terms of accretion: learning is adding to a pre-existing knowledge base, which may be organized as a set of internal symbols, or as an increasingly complex neural structure. The ecological answer is framed in terms of change: learning is change in a subcomponent of the animal-environment system which tends to make the animal gradually more adapted to some part of its surroundings ([Gibson and Gibson, 1955](#)).

The traditional view of learning is encapsulated in the doctrine of the blank slate. This says that infants are born with the ability to accumulate knowledge, but without necessarily having any knowledge built in at the outset. Much argument has been engaged in over the the question of whether the slate can really be blank at birth. To deny inborn knowledge, it is said, is to deny human nature ([Pinker, 2003](#)). I will here

⁹For a discussion of the concept of content, and a defence of the claim that content is not needed to explain behaviour (albeit mostly limited to basic organisms), see [Hutto and Myin \(2013\)](#).

suggest that the problem with the doctrine of the blank slate is not so much that the slate is blank; the problem is that it is a slate. A slate can only act as a device for storing symbols. This is not a suitable metaphor for describing a living organism that actively adapts to its surroundings.

As an alternative to the blank slate, therefore, I propose *the doctrine of the new ball*. This will be an analogy based on a well-known phenomenon in cricket. It is a recognized part of the game that the behaviour of the cricket ball changes as the ball gets older. Specifically, when bowled, an older ball can be made to swing, or curve, in a different way from a new ball. A cricket ball is made of two leather-covered hemispheres separated by a seam. On a new ball both sides are polished. As the game goes on the bowling team deliberately keep one side polished while allowing the other to wear down and become rough. The new ball can be made to swing in one direction—curving away from the batsman, say. This is called conventional swing; it is a consequence of aerodynamic forces acting on one polished side and on the seam separating the two halves of the ball (needless to say, the details are not important to the present analogy). An older ball can be turned over and bowled with the polished side facing the other way. This creates what is called reverse swing: the ball now swings in the opposite direction. ‘The whole beauty (and success) of this phenomenon is that a bowler who could bowl only outswingers at the onset (with the new ball) can now bowl inswingers *without any change in grip or bowling action*’ (Mehta, 2005).

The point of this is that it is clear that the behaviour of the ball has changed, and it is equally clear that this change is not the result of any increase in symbolic knowledge on the ball’s part. Rather, what has changed is something in the physical properties of the ball. The change is brought about here by the actions of the ball’s environment (the actions of the bowling team, and the impacts of the ball with various surfaces). This gives us a view of the process of learning that is in quite clear contrast to that provided by the blank slate. Of course, the new ball analogy is not perfect. A ball is still an inanimate object, which can only be acted upon by its environment; it is not an actively exploring, self-adapting machine like a learning animal is. Nevertheless, the contrast is perhaps clear enough to establish that we really are dealing here with two quite different views of learning: one is accumulation, the other is change.

What learning animals have that cricket balls lack is that animals stand in a par-

ticular kind of relation to their surroundings, namely a perception-action relation. In the previous section it was suggested that affordances are the properties of an animal-environment system that allow the parts to fit together. I here suggest that perception-action is best conceived as the mechanism which allows those parts to interact. It is the process by which environments change animals and by which simultaneously animals change their environments. This formulation will be a key part of the account of collaborative activity to be developed.

1.2 The radical empiricist ontology

What has been presented so far is the core of the ecological account of perception-action as it has been developed to capture the phenomenon of a single animal engaging with the world, or rather with some particular aspect of the world, one piece at a time. We have so far described perceiving as attending to particular pieces of structure, and of affordances as single correspondences standing directly between the animal and some particular physical assembly of surfaces.

But this is clearly not enough to encompass the range of behaviours that people engage in. Human behaviours that we think of as among the most primitive involve using one object to act upon a second object. We use tools to change the structure of other surfaces; we throw stones into ponds, use sticks to extend our reach, brushes to smear paint onto a wall, and so on. And when we act with other people it may be the case that what we are doing is acting directly upon them in an effort to change them in some way to our liking, but more likely we are drawing their attention to some outside object, or acting with them to change some component of our shared environment, or engaging in some other endeavour that involves not only two people but some other structure besides. All of these things go beyond the simple dyadic engagement of a perceiver-actor with a single piece of structure (be it an object or another animal). In order to incorporate these basic behaviours into our framework, then, we need to extend our ontology in some way. I have been persuaded that the right ontology for the job is that provided by radical empiricism, a philosophy which asserts not only that external objects are real and directly perceivable, but that relations between external objects are equally real and equally accessible to direct perception.

The radical empiricist perspective originates in the later writings of William James, collected in his *Essays on radical empiricism*, and was developed in the work of James' student E. B. Holt (Holt, 1912, 1915). Holt renamed the approach the New Realism. The programme failed to gain much traction outside of a small band of James' followers; this fact has been attributed to the simultaneous rise of John B. Watson's overly simplistic behaviorist programme (Charles, 2011). But perhaps another reason why the New Realism did not take off in the early part of the twentieth century was simply that it had no satisfactory account of what relations are or how they might become available to perception, and it therefore provided no suitable basis for an empirical programme in psychology. What was missing was perhaps a conception of the organism as occupying a place within an ecosystem. Heft (2013) argues that the separation of mind and body implied by internalist approaches is a consequence of 19th century psychologists' failure to adopt Darwinian ecology, which places the organism in a field of relations with its environment and with other organisms. Nowadays, however, we do have a suitable ecosystems perspective in psychology; this is what Gibson's ecological approach is all about. (Gibson was himself mentored in graduate school by E. B. Holt). There is reason to believe that this perspective can provide a useful set of tools for a study of social phenomena in the radical empiricist tradition (Charles, 2011).

Ecological psychology as a field, following Gibson, has been somewhat ambivalent about its radical empiricist roots. As a consequence, ecological psychologists have attempted to find other ways to incorporate basic triadic phenomena into their framework, usually by recourse to the concept of affordances. This is at least partly the result of some somewhat ambiguous statements in Gibson's final book. In his chapter on the theory of affordances, Gibson had this to say about social phenomena (1979, p. 128):

The other animals afford, above all, a rich and complex set of interactions, sexual, predatory, nurturing, fighting, playing, cooperating, and communicating. What other persons afford, comprises the whole realm of social significance for humans. We pay the closest attention to the optical and acoustic information that specifies what the other person is, invites, threatens, and does.

Subsequent attempts by others to expand on this have invoked the concept of affordances to deal with the perception of personality traits (McArthur and Baron, 1983), with language (Worgan and Moore, 2010), and with collaborative activity (Marsh et al.,

2006), among other things. Clearly the concept of affordances has to play an important part in our understanding of these phenomena in ecological realist terms. Nevertheless, the concept of affordances may not be the right place to start.¹⁰

Affordances are a property of the higher-order structure that exists between an experiencing animal and some aspect of its environment. In order to understand where these affordances fit in the overall scheme, and in order to work out how this description can be reconciled with the kinds of triadic phenomena mentioned above, it is necessary to begin with a rough description of the more fundamental structure that exists to support these activities, and out of which affordances arise. This description is provided by the radical empiricist ontology. The radical empiricist ontology also fits together neatly with two concepts from evolutionary ecology—the concept of the ecosystem and the concept of niche construction. Together, these ideas may constitute a suitable framework within which to build a psychology of collaborative activity.

1.2.1 The reality of relations

Since the distinguishing feature of the radical empiricist philosophy is that it posits the reality of relations, it will be as well to say something about what relations are. I will illustrate this with two examples involving dogs.

A standard example of a relation is the relation of height—of one thing being taller than another. This is a relation we might readily see when we line up two dogs and stand them next to one another. It is perhaps easy to accept that what we can see is both a pair of dogs and a relation between them (although of course the cognitivist may counter that what is seen is a pair of dogs, while the relation is inferred from the mental representation). An interesting question arises as to whether it is ever possible to see the relation without being able to see the two entities. Suppose we can only see one dog, the other being obscured behind a wall, say. Nevertheless, the relation may reveal itself in how the dog that we can see acts towards the one that is presently obscured.

¹⁰Costall (1995) provides an insightful discussion of the origins of the problems here. He shows that Gibson's thinking maintained a crucial ambiguity about the status of the social. Gibson was attempting to maintain two conflicting ideas: on the one hand, that meaning is a relational property arising from the mutuality of animal and environment (and here it is acknowledged that the environment encompasses other actors), but on the other hand, that in order to avoid slipping into cultural relativism it is necessary to assert that the asocial physical structure of the environment is primary. This ambiguity seemingly threatens to relegate the social to the status of secondary property.

The dog we can see will direct its gaze upwards or downwards in a characteristic way. Granted the information available to the observer here is ambiguous: we may round the corner and see that the second dog is in fact a small dog stood on a box, or is not a dog at all. But ambiguous or not, the structure exists from the point of view of the perceiver.

For the second example, consider a dog on a leash, being taken for a walk. The leash constitutes a physical connection between a dog and its owner. There is no need to worry about whether the leash is perceivable. Now consider a dog without a leash, running free in the park. Is it now impossible to see whose dog this is? Not at all. The connection between dog and owner reveals itself in the way the latter periodically calls out, or the way the dog periodically runs to catch up. Take away the leash and there is still something connecting dog and owner. The radical empiricist might be inclined to say that this connection is a relation, perceivable in the ongoing stream of events.

1.2.2 The animal in a field of relations

The animal is not merely an actor presented with a static environment upon which to act. Rather, the environment is itself a constantly changing thing. It is an ecosystem, made up of inanimate entities and other animals, including mates, predators, and potential prey. It is a source of diseases and of resources for food and shelter, and it is affected by the seasons and by other disruptions which may bring resource shortages or other disasters. These ecosystems are what we have evolved to live in. An ecosystem is a system of entities joined together by a network of relations (cf. [Rietveld and Kiverstein, 2014](#)).

In order to survive, an animal must not merely be able to act upon a passively perceived environment, but must be able to react to relevant changes in its surroundings. It must be useful, then, for an animal to be able to perceive the relations in its ecosystem directly, and not merely wait to be acted upon.

It was suggested above that perception-action is a relation that stands between an animal and a piece of external structure. This relation may itself provide a useful source of information for a second, observing animal. If relations are amenable to direct perception, then the observing animal must be able to directly perceive what

the first animal is doing, at least under certain conditions. Furthermore, the fact that perception is extended in time means that it does not matter if the observer cannot immediately discern what the actor is up to: this will be revealed as the action itself progresses and as the observer is able to gather more information. The suggestion here is that the perception of others is about attending to the right structure in a set of changing relations between entities, and not about inferring mental content from observed actions. This is a basic commitment of the radical empiricist account of social perception (Charles, 2011).

1.2.3 The animal as a builder of its environment

The final concept I will introduce in this chapter is the concept of niche construction. This is by now a concept fairly commonly invoked by psychologists, particularly in relation to cultural change (Laland et al., 2000). The concept comes from evolutionary biology. It was originally intended as a correction to certain dogmatic assertions that biologists have been given to make about how adaptation works (Lewontin, 1983; Laland and Sterelny, 2006). Biologists pursuing a strong adaptationist programme have held that evolutionary selection pressures are something imposed on organisms by their environments: the relationship is held to be unidirectional. Niche construction theory says that this is too simplistic. While it is true that environments do constrain what animals can do, it says, it is equally true that animals actively change their environments through the activities they engage in. And such animal-created changes can themselves feed back into selection pressures.

Some common examples are: that birds build nests, that beavers build dams, and that worms change the chemical make-up of the soil in a way that suits their digestive systems. All of these are actions that shape the physical habitat in which young members of these species live. Since these developmental environments are shaped in an important way by the actions of the species, it can no longer be true that selection pressures are simply furnished by nature. Rather, there is now a selection pressure in favour of the animals that are best suited to these very constructed habitats.

The standard human example concerns the origin of lactose tolerance in western populations. It is hypothesized that what came first was a cultural innovation: people

started keeping cattle and drinking cows' milk, and only as a consequence of this did a selection pressure arise in favour of lactose tolerant humans. This process is known as the Baldwin effect: adaptation as a result of a selection pressure arising out of the activities of previous generations (Baldwin, 1896).

A similar problem appears at the scale of psychological study, where a similar unidirectional environment-animal link has often been assumed by cognitivists. In this version, the world provides input to the cognitive system which then treats that input as a symbolic problem to be solved internally. These two ideas about environments imposing constraints on both evolution and cognition are neatly brought together in a suggestion made within cognitivist evolutionary psychology that human culture has arisen as an adaptation to the 'cognitive niche' (Pinker, 2010). The argument here is that the entire human world was in some sense already there before humans evolved: the niche already existed, it was just waiting for a species to come along and evolve into it.¹¹

But at the psychological level, too, there is no good reason why the direction of fit should be unidirectional—why the environment should provide input to the cognitive system but not vice versa. The distributed cognition perspective makes a big deal of the observation that in the carrying out of many tasks, we actively make use of the environment itself: we draw diagrams to make sense of word puzzles, physically manipulate jigsaw pieces to see where they fit, make notes of things we need to remember, and so on (Kirsh, 2010). Moreover, when we change the environment for ourselves we are also changing it for others. This is clearly a fact which should be of some relevance to any analysis of collaborative activity.

Clearly human niche construction is of a different order to the nest-building of birds or ants. Human activity has so comprehensively altered our surroundings that we are now largely unaffected by the instabilities of natural ecosystems: changes in seasons, periodic resource shortages, threats from predators, and the like. Or at least we have the impression that these things no longer affect us. The human ecosystem—the system

¹¹'The cognitive niche is [...] based on the idea that in any ecosystem, *the possibility exists* for an organism to overtake other organisms' fixed defenses by cause-and-effect reasoning and cooperative action—to deploy information and inference, rather than particular features of physics and chemistry, to extract resources from other organisms in opposition to their adaptations to protect those resources.' (emphasis added; Pinker, 2010).

of entities in our immediate surroundings—is in great part constructed by the previous activities of the species. If human cognition was all about internal reasoning, why would we go to the trouble of constructing so much stuff? A psychology of human activity must at some point address the reality and significance of material culture (Sterelny, 2014; Searle, 1995; Malafouris, 2013). The concept of niche construction will therefore be one of increasing importance to psychologists.

1.3 Summary

A basic commitment of the ecological realist approach is to the idea that many of the things that psychologists are inclined to attribute to minds are better understood as aspects of an animal-environment system. The first part of this chapter aimed to demonstrate that the ecological realist approach provides a genuinely different way of conceptualizing the animal's place in the world, compared to the way it is conceptualized within cognitivism. The ecological realist perspective does away with the traditional animal-environment dichotomy. The aim of this first part was to contest the view that ecological realism is merely a different way of describing the input, a description that in the limit maintains the old animal-environment dichotomy after all, and still requires cognitivist explanatory mechanisms (this seems to have been the conclusion drawn by David Marr). The second part of the chapter sketched some extensions to the ecological realist ontology that will be of some use as we attempt to incorporate into our scheme activities involving collaboration between multiple animals. These activities are generally triadic or polyadic: they involve not just a meeting of a single animal and a single object, but potentially of multiple animals and multiple objects at the same time. They are a test for any fundamental psychological theory.

Chapter 2

Ecological realism socialized: Acting in a populated environment

The concept of the animal-environment system is designed to capture how it is that an individual animal coordinates its behaviour relative to its surroundings. But what about situations involving multiple actors? How might these be brought into our scheme?

It is commonly assumed that collaborative activities must be treated as a special category, distinct from the kinds of phenomena that have traditionally been studied by psychologists—the kinds of phenomena that can be studied by asking individual research volunteers to solve problems in isolated booths.¹ I will set out the case, in this chapter, that this is actually the wrong way to look at things. In fact, it is unnecessary, and perhaps counterproductive, to assume that collaborative activities are inherently special and different. This is because, in reality, it is not the case that we live in a world where we have to perform activities in isolation from our peers. From birth, we are surrounded by other actors. All activity is activity that takes place in a populated environment.

To draw a line between individual behaviour and group behaviour is to create a division that does not exist in nature. In nature, the kinds of things that we are accustomed to calling ‘collaborative’ or ‘social’—playing a musical instrument in an orchestra, making movements for the benefit of one’s teammates in a soccer team, keeping to one

¹For a brief overview of some existing approaches to the study of collaborative activities, see Appendix A.

side in a corridor to allow people coming the other way to pass—are behaviours that emerge in development as the child learns to participate in ongoing activities in the world with increasing levels of skill and increasing acuteness of attention. We grow into our ecological environment and grow with them rather than constructing those environments from the ground up.

In this chapter I will restrict the discussion to non-verbal activities. In the next chapter I will turn to speaking.

2.1 Against ‘joint action’

In recent years, much of the research on action in multi-actor settings has been carried out under the label ‘joint action’ (Sebanz et al., 2006). There exists an influential line of reasoning within this ‘joint action’ literature, particularly in work by philosophers, in which an attempt is made to define strict diagnostic criteria for identifying which activities we should count as genuinely collaborative, and which we should exclude from this category.²

This way of thinking is predicated on the assumption that collaborative activity is a natural kind—the idea that all collaborative activities have something in common that distinguishes them from other kinds of activity. It is then taken to follow that in order to understand collaborative activity we first have to discover the source of this specialness. There are two good reasons to reject this. Firstly, it is impossible

²Two names often come up here, those of Michael Bratman and Margaret Gilbert. Bratman (1992) proposes three features as being characteristic of joint activities: (i) each party is mutually responsive to the other’s intentions and actions; (ii) each has an ‘appropriate commitment’ to the joint activity; (iii) each is committed to supporting the efforts of the other. These criteria may capture some of our intuitions about what ‘joint actions’ are. It is not clear, however, how such criteria could serve as a guide for telling us where to look for explanatory mechanisms. Note also that the second ‘feature’ here seems to invoke the very concept it is supposed to illuminate (although Bratman believes he can avoid this circularity by inventing the phrase ‘cooperatively neutral joint-act-type’ and appealing to that). Gilbert (1990) centres her discussion on a single ‘paradigmatic’ activity, namely, going for a walk together, which she understands in terms of rights and obligations that hold between the individuals. She argues that what makes this a true collaborative activity is explicit consent on all sides to form part of a ‘plural subject’: ‘in order to go for a walk together, each of the parties must express willingness to constitute with the other a *plural subject of the goal* that they walk along in one another’s company.’ [Her italics.] This is a distinctly legalistic description, and it is doubtful whether it describes the first-person concerns of people engaging in activities together in general. It also seems to rule out children and other animals from being able to engage in joint activities, unless we grant that they too are concerned about enforcing contractual commitments.

in practice to come up with a workable set of necessary and sufficient conditions that pick out a set of phenomena that everyone agrees is *the* set of phenomena of interest. Consider two activities: a group of ants in the process of building a mound, and two people carrying a sofa together (this is one of the go-to examples that philosophers reach for when discussing joint action). Both of these could be described as ‘joint actions’, but what could possibly unite them in a basic sense? What could be the true essence that ensures that these two activities count as ‘joint actions’, while other non-joint actions are excluded? ‘Collaborative activity’ is, in fact, just a label that we use to pick out a broad set of roughly similar phenomena, with the usual boundary conditions. It is not a natural kind.

Secondly, the view that there are genuinely such things as ‘joint actions’ goes along with the assumption not only that there *is* something special about collaborative activities, but that, whatever that something is, it is part of the causal machinery involved in the carrying out of any given collaborative task. It is assumed, for instance, that if a group of individuals do in fact successfully collaborate on some task, then each of those individuals must necessarily *know*, in some sense, that what they are doing is collaborating. But this is begging the question: it is positing that collaboration is special, and then inviting us only to look for evidence that confirms this. A more productive approach may be to assume that there is in fact *nothing* special about collaborative activity that distinguishes it in all cases from individual activity, and that we should proceed on that basis until we get to a point where the empirical facts tell us otherwise.

When we talk about ‘joint action’ we are in effect mixing two different relational categories. ‘Jointness’ describes an interpersonal relation; it denotes the fact that a group of actors can appear to be doing more than merely acting simultaneously, they can appear to be acting together. ‘Action’, on the other hand, is descriptive of a relation between an individual actor and some aspect of its environment. When we act, the action brings about change in our ecological environment. One problem here is that, at least for traditional psychological approaches, it is not immediately obvious that action *is* about animal–environment relations. Action is more often discussed in terms of the presence or absence of some mental entity or volition, as in Donald Davidson’s claim that an action must be ‘intentional under some description’ (Davidson, 1980). But on Gibson’s (1979) definition of the environment, an action *must be*

about animal–environment relations. Even an action as simple as shaking one’s head involves the environment. In order to coordinate that particular action in the first place, or to perceive whether the action has been carried out successfully, one must create the perception of optic flow while keeping static the invariant structure that specifies the layout of the environment (when you shake your head your visual field moves but the environment does not).

To talk of ‘joint action’ is to suggest a dichotomy between ‘individual action’ and ‘joint action’. The term invites an explanation of the behaviour of interest primarily in terms of interpersonal relations. Other kinds of actor–environment relations are left out of the analysis, or pushed to the background, taken for granted, described as mere ‘low-level’ action (e.g., [Tomasello, 2014](#)).

It is worth pausing over this phrase, ‘low-level action’. What could it mean? Why should actions directed at inanimate objects be considered ‘low-level’ while actions involving other animals are considered high-level or social? Why are we tempted to make this distinction?

This distinction has arisen, I suggest, as an accidental by-product of the way that action has traditionally been studied. Psychologists have tried to study behaviour by isolating individual actors from their natural environments and placing them in carefully constructed artificial environments in which the sensory inputs and outputs can be tightly controlled and measured (this is discussed at some length by [Bronfenbrenner, 1979](#)). It is tempting to believe that by getting a subject to respond to dots on a screen we are ignoring everything that is not action and perception at the most basic level.

But real environments are not like the environments created in the laboratory. An ant’s environment always contains other ants. That is to say, any action that the ant might perform is one that potentially perturbs the environment of another ant. It is this fact that allows the ants as a collective to construct great structures that could not possibly be conceived by any of the ants individually. Each ant merely acts in, and relative to, its own environment, and in doing so it alters the structure of the environment of its fellows.

‘Low-level action’, then, is an analyst’s construct. In reality, all action is social because all action has the potential to perturb the environment of another animal. The phrase I propose, ‘acting in a populated environment’, collapses the dichotomy be-

tween joint action and individual action. There is simply no need to establish a joint goal or to construct, ahead of time, a joint actor that is constituted by a set of interpersonal relations. It is sufficient for an individual to act on the available structure; this act can then be amplified or responded to by another individual whose environment has been perturbed.

The pay-off of this move, from talking about ‘joint action’ to talking about ‘acting in a populated environment’, is that it allows us to avoid having to ask the question of how it is that *the group* comes to perform some task together. Existing accounts of joint action attempt to explain such behaviour primarily in terms of interpersonal relations, asking the question: ‘How do individual actors in some situation combine to make a joint actor or team?’ (See the discussion in Appendix A). On the alternative view that I have outlined, this question is no longer seen as a pre-requisite to understanding the behaviour. Instead, we can ask a more tractable kind of question, focusing on some specific task and asking: ‘How do we identify precisely which structures in the actor’s environment are implicated in the control of the actor’s behaviour?’ (This methodological pay-off will be taken up further in chapter 4.)

2.2 Acting in a populated environment

I have suggested that there is no unifying thread that runs through all instances of the activities that we are inclined to call ‘collaborative’. In place of trying to identify general principles of collaboration, then, it will be necessary to look at some particular cases, to see what lessons they might yield. I will here briefly discuss two such cases: children’s soccer games, and pack hunting behaviour observed in chimpanzees.

2.2.1 Soccer

A famous passage from Maurice Merleau-Ponty’s writings on phenomenology describes the soccer field in terms of how it immediately appears to an experienced player of the game (Merleau-Ponty, 1963, p. 168):

For the player in action the football field is not an “object,” that is, the ideal term which can give rise to an indefinite multiplicity of perspectival views

and remain equivalent under its apparent transformations. It is pervaded with lines of force (the “yard lines”; those which demarcate the “penalty area”) and articulated in sectors (for example, the “openings” between the adversaries) which call for a certain mode of action and which initiate and guide the action as if the player were unaware of it. The field itself is not given to him, but present as the immanent term of his practical intentions; the player becomes one with it and feels the direction of the “goal,” for example, just as immediately as the vertical and the horizontal planes of his own body. It would not be sufficient to say that consciousness inhabits this milieu. At this moment consciousness is nothing other than the dialectic of milieu and action. Each maneuver undertaken by the player modifies the character of the field and establishes in it new lines of force in which the action in turn unfolds and is accomplished, again altering the phenomenal field.

This passage provides an evocative description of the kinds of structures that the experienced player must attend to and must act with reference to. [Merleau-Ponty](#) recognizes that a soccer game does not work through a series of orderly transitions of possession—the ball does not simply move from one player to another. The situation is constantly shifting, and opportunities for action are constantly arising and disappearing as the players move about. When one player passes the ball to another this is best understood not as the first player ‘losing’ something and the second ‘gaining’ the same. A pass is best understood as an action that propels the ball into a space for the teammate to run into ([Gréhaigne et al., 1997](#)). This action alters the structure of the space for the passer, and simultaneously alters the structure of the space for all of the other players.

The kinds of structures that [Merleau-Ponty](#) describes are not, however, immediately accessible to a player who is brand new to the game. When young children are first introduced to the game, they do not try to arrange themselves into organizations in order to exploit the openings between opposing players. Young children’s soccer games are characterized by the entire group chasing the ball around the field en masse. ‘A child’s basic urge is to run and chase the ball’ ([Quinn and Carr, 2006](#)). Over time, players learn to attend to information that specifies the prospective movements of the other players (dynamical structures that specify where those players are *going to be*), and they learn to exploit such structures ([Gréhaigne et al., 2001](#)). In order to reach such levels of skill, it is necessary to first learn such basic things as how the ball behaves

when you kick it, and how other players respond to the ball being propelled in a given direction. To perceive that a ball can be passed into a particular space for a teammate to run onto it, it is first necessary to be able to perceive that a ball can be passed into a particular space. All of this is learned at a very intuitive level. The child does not need to learn to play the game in some abstract, or theoretical way, using words, but learns to recognize and respond to contingencies that arise in real time, through taking part in the activity.

2.2.2 Chimpanzee pack hunting

Another example of a real-world collaborative activity is the pack hunting behaviour of some chimpanzees. Some wild groups of chimpanzees have been observed to hunt monkeys, the group encircling the prey before one of the chimpanzees is in a position to attack and kill it (Boesch and Boesch, 1989).

There is some debate about whether this is really a collaborative activity at all. Michael Tomasello (2014) has argued that it is not, and that in fact each individual chimp's behaviour can be explained by selfish motives (p. 35): 'chimpanzees in a group hunt are engaged in a kind of co-action in which each individual is pursuing his own individual goal of capturing the monkey'. Tomasello is here engaging in the philosophical game discussed above: dividing the world a priori into 'joint' and 'individual activities'.

Another description of the hunt, provided by Boesch (2002), posits the existence of 'roles' as causal components of task performance. Boesch identifies four roles: chimpanzees can act as driver, chaser, blocker, or ambusher. The suggestion is that these roles have some normative significance: a chimpanzee's role determines the share of the meat that they are entitled to. If this is the case, then some form of mind-reading must be involved, as each chimp would need to know what role the others are playing.

But is it in fact necessary to suppose that the chimpanzees are really taking on some well-defined 'role' in these situations (even if we acknowledge that such a role might be implicit, and not represented in a language of thought)?

Tomasello (2014) describes the hunt in terms of individual goals; Boesch (2002), in terms of individual roles in the context of a single shared goal. What both of these

description schemes obscure is the importance of external space relative to the immediate positions of individual hunters and prey. In a successful hunt, individual hunters will surround the prey in such a way that all potential escape routes are blocked. Muro et al. (2011) present a computer simulation of a similar type of hunting behaviour in wolves. In the wild, wolves are observed chasing a buffalo until the buffalo becomes tired and collapses, at which point the wolves surround their prey on all sides. Muro et al. were able to reproduce this behaviour by modeling the wolves as agents following two rules: 1) get as close as possible to the buffalo without putting yourself directly in its path (i.e., avoid being trampled on), and 2) maximize your distance from neighbouring wolves. It is tempting to conclude from this that wolves need not possess any kind of explicit understanding of the ‘role’ they are occupying, but rather that they are literally following a couple of simple rules. But perhaps even these rules are unnecessary outside of the constraints of the computer simulation: in reality, rich perceptual information is available to the hunter in the unfolding structure. The question is, what aspect of this structure is it that hunters must learn to respond to?

This is the question that concerns the radical empiricist. A plausible answer is that hunters are looking for the biggest gap in the space surrounding the prey that isn’t already occupied by a fellow hunter. That is, individuals are responding to *spaces*, not to distances from other animals per se: there is no need to calculate the distance from one’s neighbours and deliberately adjust one’s position in order to maximize this distance, nor is there a need for any kind of internal bird’s-eye schema of the overall situation; rather, the apparent division of labour in the hunt simply emerges out of individuals responding to local affordances, to directly perceivable gaps.³

The skills involved in this group hunting behaviour will necessarily have to be learned to a great extent—through participation in the hunt itself and through such activities as play with other chimpanzees. Identifying exactly how it is that an individual learns to participate appropriately in the hunt is an empirical question, and one that will, in practice, be difficult to study, given that these hunts may take place over an

³A similar process may be at work when groups of people gather together spontaneously into a circle to have a conversation. When we do this we tend to arrange ourselves in a symmetrical fashion, with a regular distance between participants, or in an unsymmetrical fashion that allows all the participants to attend simultaneously to some external object in the distance. Adam Kendon (2010) calls such configurations ‘F-formations’, and suggests that these formations allow participants to maintain order in the conversation and to jointly manage one another’s attention behaviour.

extended area within a dense forest. Considering this pack hunting activity in terms of structure available to perception from the perspective of an individual chimpanzee does, however, promise to lead to a rich understanding of the mechanics of this behaviour.

2.3 Growing into one's environment: lessons from Holt and Vygotsky

What is clear, from the two cases discussed above, is that participating in collaborative activities involves two things: it involves a pre-existing environment which provides structure and provides a context for the actor's learning, and it involves the learning of skills which are necessary if the actor is to take control of the environment that it finds itself in. Learning to participate in such activities is thus a dialectical process: the environment provides structure which the developing animal learns to respond to, and in mastering these responses the animal learns to alter the layout of its environment in reproducible ways, and this in turn opens up the possibility for the actor to discover new aspects of the structure of its environment. Two authors in particular have written insightfully on this type of process: E. B. Holt, whom we encountered already in chapter 1, and Lev Vygotsky, whose career began just a few years after Holt introduced his new realism, and ended just a few years after that.⁴

2.3.1 Holt and the recession of the stimulus

Holt spoke of a process which he labelled, somewhat misleadingly, 'the recession of the stimulus' (Holt, 1915, p. 75). I say misleading because the phrase would seem, on its surface, to imply that, through the process of learning, the animal retreats into itself. In fact, Holt intends the phrase to mean precisely the opposite of this. He intends it to denote how, in learning to master basic level skills in the context of an environment, the animal becomes able to engage with the environment at ever-increasing levels of

⁴I am not aware of any evidence that Vygotsky was familiar with Holt's work. There are interesting parallels. Both were very much influenced by Darwinian thinking, which itself takes a dialectical view of the organism in an ecological context (Heft, 2014). Holt appears to inherit this most importantly from William James (1912), while Vygotsky inherits his dialectical thinking also via Marx and Engels.

structure. Whereas Holt talks of ‘recession’, he might just as well have spoken of ‘growth’, or ‘adaptation’, or ‘engagement’ between animal and environment.

Perhaps the clearest exposition that Holt gives of the concept is in the following passage (Holt, 1915, p. 186):

When one first learned to walk, the process involved lively consciousness of pressure on the soles, and at different intensities in the two feet; of visible objects which one carefully watched in order to steady oneself, etc., etc. One now walks with head in air and with almost total oblivion of the steadying visual objects and the unfeeling tactual objects with sharp corners, the stairs and the inclines, which it was once so wise to keep in view. At first one stepped, and each step was an adventure in itself; now one *walks*, or perhaps not consciously even this; for one may consciously not be walking or running, but catching a train, thinking over a lecture, bracing oneself to do a sharp stroke of business. The walking behavior, although no less behavior and no less involving functional adjustment to the environment and hence no less involving ‘content,’ has now been taken up (along with other behavior systems) and made component of a more highly integrated and elaborate form of behavior.

It is clear, from this passage, that when Holt talks of ‘recession’ he is talking about the same sort of thing as was described above, in the case of the children chasing the soccer ball around the pitch. In learning to walk, the child must initially attend carefully to the most basic adjustments in position. Later, the child is able to assemble itself into a walking system in a highly efficient manner, which frees up attentional resources to further explore the structure of the surroundings. and perhaps newly created structures such as optic-flow in the context of locomotion. Similarly, the child, when first learning to kick a soccer ball to a teammate, must first attend carefully to how different types of leg-swinging action cause the ball to be propelled in different ways. Later, when the child can kick the ball efficiently, and in a somewhat reliable manner in some roughly desired direction, this frees up the child’s attention to discover new patterns of movement that can be responded to. Now the child can start to attend not only to where the ball is going to go, but also to where her teammates are, or where they are going to be.

Holt describes the process of learning in terms of the integration of reflex arcs within the learner’s nervous system. At one point he discusses a hypothetical example of a micro-organism which learn to move towards light while also learning to back

away from a source of heat that becomes too intense. It is clear that he understands this learning process in dialectical terms, the animal and its environment together constituting the system (Holt, 1915, p. 75):

And one could not describe what the animal *as a whole* is *doing* in terms of the immediate stimuli; but this can be described only in terms of the environing *objects* toward which the animal's response is directed. This is precisely the distinction between reflex action and specific response or behavior. As the number of of component reflexes involved increases, the immediate stimulus itself [light] recedes further and further from view as the significant factor.

The recession of the stimulus is the growth of the animal into its environment. It is the result of change in the actor in accordance with the doctrine of the new ball: change that impacts the adaptedness—the fit—that exists between an animal and its environment.

2.3.2 Vygotsky and the zone of proximal development

Lev Vygotsky's thinking about learning is strikingly similar to Holt's in its dialectical nature. Vygotsky formulated a language for describing learning in a social context, a language which has been particularly useful and influential in the context of education (Lake, 2012). An important concept here is that of the zone of proximal development. In introducing this concept, Vygotsky (1978) describes the standardized testing of his day. He notes that tests administered to children were designed to measure the child's 'actual developmental level': the tests measured the children's ability to solve problems on their own, in isolation. (Of course, this remains true of much standardized testing in childhood education settings today.) Vygotsky suspected that such testing may be measuring the children's true abilities in a very restricted and limited way. He complains that the psychologists who drew up these tests 'never entertained the notion that what children can do with the assistance of others might be in some sense even more indicative of their mental development than what they can do alone' (p. 85).

The 'zone of proximal development' is intended to denote the space that exists between the things the child can do in isolation (the 'actual developmental level') and the things the child can do with the help of others. Vygotsky recognizes the latter

as important. He claims that the things that a child is able to do with help today are indicative of the child's capacity to master those things for him- or herself in the immediate future. [Vygotsky](#) noted, in effect, that private, mental activities are really a special case of actions carried out first through situated action in the context of others (a process he called 'internalization'). [Lake \(2012, p. 49\)](#) quotes a useful passage from a book by [Bodrova and Leong \(1996, p. 35\)](#), which parses word by word the phrase 'zone of proximal development':

Vygotsky chose the word *zone* because he conceived of development not as a point on a scale, but as a continuum of behaviors or degrees of maturation. By describing the zone as *proximal* (next to, close to), he meant that the zone is limited by those behaviors that will develop in the near future. Proximal refers not to all the possible behaviors that will eventually emerge, but to those closest to emergence at any given time.

In formulating the concept of the zone of proximal development, Vygotsky was particularly concerned with understanding the emergence of the 'higher psychological processes', such as language, mathematics, and scientific reasoning. I will touch on these briefly in the next chapter. But I will note here that if the existence of a supportive social context is important in learning such things as mathematics and reasoning, it is an absolutely constitutive part of the kinds of intelligence required to become a skilled soccer player, or pack hunting chimpanzee.

A child playing with a soccer ball in isolation from her peers may learn to manoeuvre the ball with great skill. She may learn to keep it up in the air, to step over it in elegant ways, to balance it on her neck. But such skills will not in themselves make her a skilful player of the game of soccer. There are skills that are impossible to learn in isolation: attending to the prospective movements of the other players; learning to recognize opportunities to pass the ball to a teammate in an advantageous position; learning to move without the ball in such a way as to distort the opponents' defensive system; learning to pass the ball back and forth with a team mate quickly, in a manner that wins territory and advances the ball towards the opponents' goal. Such skills are recognized by soccer commentators as hallmarks of intelligent play. But this is a kind of intelligence which, as in [Vygotsky's](#) day, is not typically measured by standardized testing.

2.4 Summary

This chapter has made two major arguments. Firstly, there is no special problem of ‘joint action’, and there is no sharp divide, from the perspective of the actor, between engaging in actions that are ‘joint’ and engaging in individual action. We are born into a world that contains other actors, and any action we engage in is therefore an action that takes place *in a populated environment*. Some of our actions may directly perturb another actor; other actions may alter the layout of the world in a way that impacts on how the environment appears to other actors. In rejecting the dichotomy between the joint and the individual, we free ourselves from such problems as explaining how it is that we come to coordinate with ‘other minds’, or how it is that a group is able to emerge as an agent. Such concepts are not necessary. Collaborative action is possible because multiple animal-environment systems co-exist in a single world, and the actions of any one animal can thus immediately impact the activity of another.

The second point is that learning to engage with others collaboratively is a dialectical process—one in which the learner first attends to basic structures in the environment, and in responding to these opens up new opportunities to learn. This is true for the non-verbal activities that have been discussed in this chapter, and it is also true of those functions that [Vygotsky](#) described as the ‘higher mental processes’, the most central of which is language—the subject of the next chapter.

Chapter 3

An ecological realist guide to speaking

In psycholinguistic theorizing, speaking—articulation—is said to be the end product of a chain of events that starts with a thing that a speaker wants to say, which is initially encapsulated in a mental concept or lemma; this lemma feeds into phases of word selection and motor planning before finally being realized as an utterance (Levelt et al., 1999).¹ What is this lemma, in its initial state? Clearly it is supposed to be a kind of Brentano-style intentional object: a mental counterpart of the thing that eventually gets realized as movements of the vocal apparatus. But why suppose that such a thing exists? A powerful reason for thinking so comes from our own experience. When we think to ourselves we find that we are constantly formulating sentences in our heads. This gives us the impression that what we choose to utter out loud is a subset of the sentences of this ‘inner speech’. We thus have a strong intuitive reason for suspecting that all utterances must be preceded by corresponding acts of thought. And if all individual utterances are preceded by an internal thought, then it is plausible to imagine that thinking is a primary phenomenon that must take basically the same format in pre-verbal infants as it does in adults. That is, just as a thought precedes an utterance, the ability to speak (in general) is likewise preceded by an ability to think (in general). On this view, thinking happens in a language of thought, and language learning is the

¹A much-edited version of this chapter appears in the journal *Ecological Psychology* with the title ‘A radical empiricist theory of speaking: Linguistic meaning without conventions’ (Baggs, 2015).

process of mapping this pre-existing thinking system onto whatever natural language happens to be spoken in the vicinity of the infant. And the same may be true for other animals as well: thinking happens in mentalese; it is just that no other species happens to be capable of mapping its thoughts onto an external language in the way a human infant can.

So one view of things here says that spoken utterances are externalized thought (Chomsky, 2011). But the reverse may instead be true. It may be that speaking out loud is primary, and that the ability to formulate internal sentences in inner speech—the ability to speak in one’s head—is in fact a consequence of having mastered the out-loud variety (Vygotsky, 1962, 1978). On this alternative view, the utterances of an infant at an early stage of language learning need not be the product of an internal thought at all. Rather, these utterances may simply be an instance of the child attempting to act upon its environment directly. That is, utterances are exploratory actions. Only later, after extensive exploration—after the child has tested out a wide variety of speech-actions and learned their meanings by attending to the things that these actions bring about—does it become possible to bundle a selection of utterances together into something like a ‘thought’, where a thought is really a sentence in inner speech. This view of things is perhaps less intuitively appealing than the first, at least from the standpoint of a linguistically sophisticated adult human. But it is a view that can perhaps more easily be reconciled with the fact that language has to be learned. (And incidentally it seems to suggest an explanation for the fact that adults feel the need to tell children to ‘think before you speak’.)

Of course, the present realist framework does not admit representations into its description scheme: internal mental objects are excluded as a matter of principle. The first view of things—that spoken language is representational thought externalized—is thus unavailable to us. The problem for the ecological realist is how to characterize language without invoking mental terminology, and in a way that is consistent with the tenets of ecological realism: that the things we have access to in experience must be accessible to perception.

Ecological psychologists have attempted a number of schemes for reconciling language and ecological realism, none of which is entirely satisfying. These attempts have generally been of two types. One type takes some existing concept from Gibson’s

theory—affordances, energy arrays, information, event perception—and attempts to repurpose it for language. The other type tries to co-opt the existing mentalist description scheme we use to talk about language, and to strip it of all its dualistic attributes. I will discuss some of these proposals below before outlining an alternative based on the radical empiricist ontology. But first it will be necessary to establish what the problem space is.

3.1 Does ‘language’ exist?

The first issue that must be addressed is the question of what kind of thing we are talking about when we talk about language. On the view of language as externalized thought, this appears to have a fixed answer: language is a system for arranging ideas. On the alternative action-based view, the issue is far less clear. One seductive answer is that language is a kind of medium: speaking is a kind of acting, and it is performed through the medium of language. I believe this view is quite wrong. An argument against the language-as-medium concept can be made on much the same lines as William James’s argument against the existence of ‘consciousness’ ([James, 1904](#)).

The problem with the idea of consciousness is not the fact of subjective experience, but the implication that experience goes on inside some kind of container. If consciousness is a container, then there arises the familiar problem of how things on the inside of it (thoughts) can ever interact with things on the outside (objects). What James rejects is ‘consciousness’ as a label for this container: ‘It is the name of a nonentity, and has no right to a place among first principles.’ In short, the rejection of consciousness is a rejection of mind-body dualism. James recognizes, however, that consciousness is also invoked to name a function—the function by which an individual comes to know things. He proposes to re-conceive this function as a direct relation, ‘knowing’, that exists between an experiencing animal and an object.²

The parallel argument for ‘language’ must reject a different kind of dualism: language-environment dualism. A common way of talking about language, or communication, is to say that a message is transmitted from a speaker to a hearer through a ‘channel’,

²This general idea was later taken up by Gibson and refined into the concept of perception-action; see [Heft \(2001\)](#).

i.e. a container, or rather a conduit (Reddy, 1979). It is easy to see how this description could be applied to a telephone line or an internet chat interface: there, something is actually being encoded at one end and transmitted through a physical network of wires to a recipient at the other end. But the concept of the channel is freely applied by theorists to describe speaking in the normal case: between two or more people in the same space, separated only by air. This is odd because nobody has ever seen a speech channel connecting two people talking in this manner. The obvious candidate for what constitutes the channel is the collection of sound waves that are produced by a speaker and attended to by a listener. So perhaps that is why nobody has seen any such channel: the channel is invisible. But then again, what type of existence could this channel have? Does it still exist when two people are in the same space but not talking to one another, or does it come into existence only for the duration of the relevant sound waves? And if perceiving someone's speech is really attending to structure in sound, why should that sound need to be sent through any kind of conduit at all? Why can it not just be out there?

The problem with the conduit metaphor is best seen if we consider an interaction between two people talking about something in their immediate surroundings. What the conduit metaphor does is it distracts us from the reality of a shared world: it creates the impression that in order to engage in an activity together we must somehow establish that this is the activity we are engaging in, and that we must do so by sending messages back and forth down the pipe. Any further verbal utterances that are made while the activity is ongoing are also conceptualized as messages sent through the pipe. The verbal actions are seen as not exactly *part of* the activity; they exist as a kind of commentary on how the activity is going, or as a secondary activity that interacts with the primary activity somehow. There is a division between the realm of physical action and the realm of language. But if we consider even briefly the kinds of actions that do in fact take place, it soon becomes apparent that this division is hard to maintain. The kinds of actions that exist do not divide naturally into just two types: verbal actions and physical actions, or communicative actions and non-communicative ones. Instead, there are physical actions that seem to be communicative (being handed a tea towel by someone who is washing up); there are verbal actions that seem to be ends in themselves (performatives); there are vocal cries that serve no obvious communicative

purpose; and there are a great many utterances that seem to be uninterpretable except as a combination of the physical and the linguistic, such as the expression, ‘What’s that!’

As with consciousness and physical reality, then, there is a case to be made that the division between language and the environment is unsustainable. But again, we must recognize that the conduit metaphor is invoked in the service of a function. What is the function of language? A plausible answer is that it is a tool for directing attention (Reed, 1995; Tomasello, 1999). The rest of this chapter will therefore be an attempt to outline how this function can be sustained in a realist framework. The suggestion will be that the action of directing others’ attention is achieved through actions performed on the system of relations that constitute the local ecosystem. If the function can be achieved through the system of relations, then there is no need to invoke any additional channel of communication.

I will here deal with a likely objection to the above. If language does not exist, how can it be that there are different *languages*? The response is that obviously there are different languages, but that these exist only in opposition to one another. To say that languages A and B are different is really only a description of a state of affairs: it is shorthand for something like the statement that, ‘When I hear those people two people talking to each other, I cannot draw any useful information out of their speech.’ But it is not the case that two people who share no common language are thereby occupying two incompatible worlds, or that they are stuck in a shared physical environment with no channel to communicate with. Two people with no common language can still point to things and produce names; this fact is the basis of much anthropological fieldwork. These two people find themselves in the same world. What is missing is not a channel, a ‘language’, but the skills to carry out certain kinds of actions; these two people are simply not (yet) built in such a way that they can point each other to bits of their shared environment in an efficient way.

3.2 Reconciling speaking and ecological realism

One of the most common criticisms of the ecological approach is that it can deal with simple things but not complex things. It can characterize simple, ‘low-level’

activities—perception-action phenomena. But because it rejects representations it has no hope of characterizing more complex cognitive activities that surely recruit memories and internal symbols. Language is the prototypical ‘higher-level’ cognitive activity. As such, it is unsurprisingly the case that a number of ecologically-inclined psychologists have attempted to take up the challenge of providing a non-representational account of language. I do not think a completely satisfactory account has yet been given. Nevertheless, it is instructive to look at the strengths and weaknesses of the proposals that have been made.

The first proposal is that perceiving someone else speaking is perceiving affordances—it is perceiving what the other person affords. In a trivial sense this has to be true: according to the framework, all perceiving is perceiving affordances. The trouble is that it is not clear how a given instance of speech could itself afford anything. The structure of speech is simply not the same kind of thing as the structure that is perceived in light, say. Light is structured in a lawful way by the fact that it is reflected off the surfaces of the environment—the structure in light *specifies* those surfaces as a consequence of the way it is reflected. Affordances, as defined in chapter 1, arise from correspondences between animals and the things in the world that those animals can act upon; correspondences that can be perceived in the structure in light. Speech structure, by contrast, is apparently arbitrary. Speech structure is characterized by what linguists call ‘duality of patterning’: a finite set of sounds (phonemes) is combined into a set of meaningful units (words or morphemes), and these meaningful units are then further combined into larger meaningful patterns or sentences (Hockett, 1960). If the structure in speech is not directly informative about external structure, then it does not appear reasonable to say that hearers perceive affordances via speech in anything like the way that they perceive affordances via light. Proponents of language-as-affordances are aware of this problem, and attempt to resolve it by talking about ‘social affordances’ (Kono, 2009) or ‘interaction affordances’ (Worgan and Moore, 2010). These terms appear in essence to be placeholders for a hoped-for solution.

Another proposal says that language is a kind of energy array. Recall that in the ecological jargon an energy array is a medium of perception. For example, the optic array is the set of all potential points at which an animal can sample the richly-structured light that specifies its environment; the acoustic array is the set of positions from which

an animal can sample the noises of its environment. The proposal is that language exists in ‘dialogical arrays’ (Hodges, 2009). This is based on the observation that, by talking to one another, conversation participants can perceive, at second-hand, information about objects or events witnessed by others, and can learn about other people’s dispositions and intentions.³ This, again, is clearly a true description in some sense. But the suggestion that language is a discrete kind of energy is puzzling. What kind of energy could it be? Clearly it is not like light or sound: dialogical energy is not just out there constantly in the environment, but has to be created by a speaker. The only plausible explanation appears to be that language is a kind of private energy that resides inside people, and that has to be goaded out through verbal actions. But this suggests a representational account (it is reminiscent of the transactive memory scheme proposed by Wegner, 1986, according to which language is a device for transferring memories from one mind to another). If this is so, the language-as-energy theory has not succeeded in providing a non-representational account of speaking.

A third proposal, from Fowler (1986), says that perceiving speech is perceiving events, where the word ‘events’ is to be understood in the ecological fashion as denoting temporally extended, directly observable disruptions in an energy array. This proposal is presented specifically as an explanation of how the *sounds* of speech are perceived, and relies on maintaining the standard division between ‘language’ and ‘context’. Fowler acknowledges that this division may not be perfect: ‘In ordinary settings in which communication takes place this is almost certainly not a natural partitioning because it leaves out several aspects of the setting that contribute interactively with the linguistic utterance itself to the communication. [...] For the present, however, I will preserve the partitioning [...]’ She then finds it necessary to divide the event of talking into two: a speech event, denoting the sound pattern itself, and a linguistic event, which is where meaning is supposed to be located. In effect, she is committed to the idea that information has to be extracted from a message—that is, to the idea that ‘language’ exists as a conduit through which ideas travel. As a tentative account of what is required for a listener to perceive the linguistic event, Fowler suggests that words and sentences are meaningful because the same structures have

³This proposal is the most closely informed by James Gibson’s own notes on language, mentioned in the preface (Gibson, 1982).

been used in the past for similar purposes: ‘listeners apprehend the linguistic work that the phonetically-structured vocal tract activity is doing by virtue of their sensitivity to the historical and social context of constraint in which the activity is performed.’ This looks very similar to more recent cognitivist proposals that language perception requires enormous statistical processing of the input (e.g., [Saffran et al., 1996](#)). How might this be achieved without representations? Clearly this is an important gap in the proposal.

A final proposal that I will discuss here says that perceiving the meaning of speech is just perceiving information. This, again, is certainly true in some sense or other. The problem is to explain the nature of this information: what is it information *about*? Spoken structure is a property of the environment, but it is not simply given in the way that structure that specifies the colour of the sky is given; spoken structure must be created by the actions of a speaker. But speakers are unstable objects, constantly moving about, and frequently engaging in different activities from one moment to the next. This being so, how can it be that spoken structure is at the same time sufficiently stable in nature that children can learn what it means?

An answer that has been frequently proposed is that words and other linguistic segments derive their meaning from *convention*. I will devote some space to this idea, because it is one that is both attractive on its face and seemingly very difficult to do without. I also think it is entirely wrong, and a hindrance to characterizing speaking in a genuinely anti-representational way. The convention-based view is explicitly endorsed by [Wilson and Golonka \(2013\)](#):

The only difference between perceptual information [for vision] and linguistic information is in the relationship between the structure in the energy array and the meaning of the information. For language, the structure in the energy array is not about the dynamics of, say, articulation; it’s about whatever the words mean. The structure comes to have this meaning because of the social conventions of the language environment and what we learn is, therefore, a conventional meaning of the pattern. This conventional underpinning gives stability to linguistic information, but the difference between a law [e.g., the law that specifies the structure in light] and a convention is very important. Conventions can change and so can the meaning of words; language is much less stable than perception. This decreased stability is, of course, a fact of language to be explained, so perhaps it is not a disaster for the analogy we are developing here.

The position, then, is that information can acquire its meaning in at least two ways: 1) through a *specifying* relationship that holds between structure in energy and structure in the environment (as is the case in visual perception); and 2) through a *conventional* relationship that holds between structure in particular (sound) patterns on the one hand, and particular entities, events, or action opportunities in the environment on the other. In support of this position, Wilson and Golonka cite Barwise and Perry's (1983) situation semantics, and Chemero's (2009) endorsement of it. Situation semantics is a formal framework within which conventions are defined as constraints which relate two different situations or events: a speech event (a sentence: 'it's raining', say), and a second event which the speech event is about (a physical event in the world: the presence of falling water). As long as a language user has access to the constraint and to one of the events, they can use these to access the second event: a listener can derive the meaning from the speech event, and a speaker, on observing a physical event, knows what speech event to produce.

But what are these constraints? They appear to instantiate a particular version of the language-as-conduit model: the idea that linguistic structure, by nature, possesses a *literal meaning*. The literal meaning of a sentence is the meaning that the sentence would have in a 'null context' (Searle, 1978). Speaking is held to be a process of constructing a sentence whose literal meaning encodes the thing the speaker wants to say. If semantic constraints are rigid relations between two events, as proposed, then conventional meaning is a species of literal meaning, and speech events are vehicles for literal meaning.

The notion that pieces of linguistic structure have a literal meaning has been very widely assumed by linguists and philosophers. A problem with it is that it does not appear to be sufficiently flexible to explain actual language use: people create novel pieces of linguistic structure all the time; how can listeners have access to the literal meaning of sentences that have never been uttered before? However, the main problem for present purposes is that literal meaning does not appear to be compatible with the anti-representationalism of ecological realism. The suggestion is that a listener, upon attending to a speech event, has access to a set of literal meanings about some ideal physical event, and that this meaning then has to be reconciled with present circumstances. The listener is again faced with a problem which seems to demand internal

processing.

Could it be possible to define conventions in such a way as to avoid invoking literal meaning? It is not clear how it could be. The problem is perhaps that conventions are things that we can only really define from outside the system, as observers looking in. We can identify regularities in the way particular linguistic structures are employed to talk about particular sets of things, and we are tempted to make sense of those regularities by grouping all of these events based on what they have in common. Thus, we are confident that we have collected enough instances of the structure ‘it’s raining’ that we are justified in claiming that a conventional relationship exists between that structure and a state of affairs in the world. But this is really a simplification of a very rich data set: a simplification because it attempts to extract the essence of a phrase, its literal meaning, from all of the supporting structure that goes with any particular instance of the phrase in action.

This is really the core problem with the language-as-conduit model: it treats language as an immutable set of objects—a set of things we can identify. This point has been invoked in support of a general criticism of the modern programme in linguistics, made most forcibly by the philosopher Roy Harris who talks of the ‘language myth’ (Harris, 1981). The ‘myth’ is that using language is putting a set of objects into action. The argument is that the way linguists think is heavily influenced by the existence of things like dictionaries. Dictionaries exist to give a description of a set of regularities that can be identified in the way particular units of linguistic structure are already used. A dictionary is a kind of simplifying descriptive device. Linguists have taken the idea of the dictionary and posited that a similar device must exist in the heads of speakers—that speakers must possess an internal lexicon, which is consulted in some fashion during normal speaking. Similarly, the structure at a higher level is abstracted and simplified in the form of syntax trees—another descriptive device. And the existence of the alphabet gives rise to the idea that the smallest units in which we perceive language are phonemes.

There is no reason in principle why the theory of language should so closely resemble these tools that have been invented, for the most part, to facilitate the *use* of (written) language. In fact, if we begin our analysis of speaking with the behaviour itself, it is quite clear that these objects—dictionaries, alphabets, and the like—*must*

have been derived from a much richer set of structure that is created by speakers. Speaking is the primary phenomenon, ‘language’ is derived.⁴ There is a parallel here with the problem of conceptualizing visual perception as image-based (discussed in chapter 1). Photographs are impoverished, static samples of a much richer array. Similarly, a dictionary entry is an impoverished sample of the things that a particular piece of linguistic structure can be used for. In both cases, the metaphor leads to a view that the perceiver is confronted with impoverished data which demands to be filled in—to be placed into ‘context’: images have to be built up into a mental model, and words have to be reconciled with knowledge about the ongoing situation. The suggestion is that the apparent poverty of the stimulus, in both cases, is an illusion created by the attempt to reconstruct a rich activity on the model of an impoverished sample of that same activity. Another way that this argument has been put is that speaking is a first-order phenomenon, a behaviour, and that ‘language’ is a second-order phenomenon, a by-product (Thibault, 2011). The first is a verb, the second is a noun. The point is that a *psychology* of language use should be concerned with the first order behaviour—the verb, ‘speaking’.⁵

The question is, how do we conceptualize the first-order behaviour of speaking in a way that recognizes its inherent richness? How do we avoid attempting to fit the first-order phenomenon into the framework provided by the existing second-order description scheme. What is needed is a different account of the machinery that underlies an act of speaking.

3.3 Speaking as acting on the relational field

A common observation is that a language is like a set of tools (Everett, 2012). This is meant to stand in opposition to the view held by certain philosophers that language is in

⁴Harris (1988) emphasizes the distinction made by Saussure between *parole* and *langue*: the former is the primary phenomenon—speaking; the latter denotes a formalized description of an underlying system that is assumed to be necessary in order to generate individual acts of speaking. Saussure proposed that linguistics should be the study of this idealized system, *langue*.

⁵Some writers have advocated the use of the word ‘*linguaging*’ as a verb to denote the first-order behaviour. I avoid this neologism, partly because I find the word ugly, but also because I think it subtly invites back in the very model it is trying to replace: how else could one engage in ‘*linguaging*’ except by using bits of ‘*language*’? I prefer the term ‘*speaking*’, employed in a broad sense intended to include the use of sign language as well as gestures and potentially other tools and technologies.

fact a self-contained system of mathematical logic. The language-as-tool metaphor is proposed in order to draw attention to the fact that when people use linguistic structure, they are using it in service of some particular function: language is used to do things. For present purposes, I assume that the functional, tool-based view is valid. But if language is a tool, what is it a tool for? A drill is a tool for making holes in a wall. This works because the drill bit comes into physical contact with the wall's surface. A sentence might be a tool for directing a listener's attention to a thing in a certain way. But what does the sentence make contact with? The obvious response is that it makes contact with a listener, who then extracts the meaning and works out where their attention is being directed to. But this is again to invoke the old mind-body dualism and to demand an explanation in terms of internal representations. The proposal here will be that what a sentence makes contact with is not the listener per se, but a *relation* that connects a listener to some aspect of the structure of the world.

It will be necessary to lay out with some care exactly what is being proposed here. Recall that in chapter 1 it was asserted that the animal occupies a place in an ecosystem made up of objects and surfaces and other animals, and of relations between all of these entities. In effect, the perceiving-acting animal stands at the centre of its own web of relations connecting it to everything else around it. The specific relations that it perceives when it explores its environment are of the affordance type. The animal, by virtue of the things it has learned to do, is able to perceive a rich set of affordances when it moves about in its surroundings.

What is now being proposed is that: a) these affordance relations exist even when the animal is not attending to them; b) these relations are themselves public, and can in principle be perceived by an observer; c) in the case of language-using humans, these relations can be *directly acted upon* by other speakers; and d) the tool for acting upon another person's web of relations is linguistic structure.

This is not to deny that speaking involves sound patterns and ears. Clearly one's ability to act upon someone else's relational web—whatever that means—must depend crucially upon the auditory perceptual system of a cooperating addressee. There can be no speaking except with a linguistically-competent addressee who possesses both a functioning pair of ears and an appropriately configured nervous system. Whatever it is that the speaker is doing to the relational web when speaking, the action of speak-

ing must also involve acting upon the machinery inside the addressee somehow. This is true, but it is not incompatible with the proposal here. To invoke nervous systems and hearing organs is to construct a description at the biological level of analysis. The present proposal is that, *at the psychological level of analysis*—at the ecological scale—what is being acted upon is the set of relations connecting a listener to other things in the speaker's surroundings. From the point of view of the speaker, speaking is not merely acting upon the addressee, but upon the relation of the addressee to something else in the speaker's environment. Similarly, from the point of view of the addressee, it is not the case that the information in a piece of linguistic structure is extracted and then reconciled with the environment. Rather, the information is perceived as *part of the environment*. Perceiving the information in the spoken structure is the same as perceiving the corresponding relation.

In order for a verbal action to provoke an appropriate response, the action must be compatible in some way with the relation it is aimed at. The action and the relation must fit together. Below I propose that this is possible because the relation is shaped by the learning history of the addressee. As a child learns a language, its relations become increasingly compatible with verbal structures produced by others, and thus increasingly accessible to actions from other speakers. The point for now is that if a verbal action can be directly compatible with a certain kind of relation, then there is no need to invoke conventional meaning or any other form of the language-as-conduit model. Both the action and the relation are real entities: they fit together in the same way as any other action produced by an animal fits together with a suitable affordance—in the same way that giraffes fit together with tall trees, or that the wings of certain birds fit together with currents of warm air.

The difficulty is that relations are not things we think of as entities that can be acted upon. They are not things that we typically name, and thus it is hard to accept that there is really anything there that can come into contact with other things or that can be implicated in the causal structure of events. But something along these lines must be necessary if we want to get beyond the second-order description scheme that encourages an understanding of speaking as an applied use of dictionaries and other artifacts. A theory of the first-order phenomenon will require an ontology that can incorporate verbal actions as discrete events, rather than as composite entities made

up of structure and context. And if what is sought is a genuinely non-representational account of language, as is the goal here, then the only option may be to view speech as actions upon the relational field. Considering all of the proposals discussed above, the conclusion is this: either relations are things that can be directly acted upon, or no plausible candidate for a non-representational account of language has yet been identified.⁶

3.3.1 Action controlled with reference to relations

It will be useful here to illustrate more fully what is meant by the idea of acting on a relation. A more accurate description is to say that the action is being controlled *with reference to* a relation. Relations are traditionally treated as a kind of property, in the same way as the colour of an object is treated as a property of that object. Relations are said to differ from things like colours in that they are not properties of a single object, but exist *between* individual objects. But in general, properties are things that simply exist. We can change the colour of an object by painting it, but it is odd to say that we are thereby acting on the object's colour property. Instead, we are inclined to say that we are performing an action upon the object itself—the action on the object changes one or more of the object's properties. What, then, does it mean to say that relations, unlike colours, can be acted upon?

Chapter 1 adopted the view that affordances are relations that stand between an experiencing animal and a piece of structure in the environment (Chemero, 2009). On this account, acting on affordances is necessarily acting with reference to a relation. An action is an event that changes the nature of the relation between animal and object. This description can readily be applied to standard examples of actions on affordances: climbing a step (Warren, 1984) is changing the step's affordance for the climber from climb-up-able to climb-down-able; throwing a rock (Bingham et al., 1989; Zhu and

⁶For completeness, I will here mention another proposed account of language which also emphasizes the importance of relations: relational frame theory (Hayes et al., 2001; Tonneau, 2004). This is an intellectual descendent of B. F. Skinner's behaviour analysis programme. An important difference between this account and the one presently being outlined is that the relational frame theory does not conceive of relations as things in the environment that can be directly perceived, but as things that a person *does*: 'People *frame events relationally* in the moment as an active process that is a function of their extensive learning history and stimulation in the present environment. "Storage" of these frames as structures is not implied and not required.' (Blackledge, 2003).

Bingham, 2008) changes the rock from an object that affords throwing to one that affords watching in flight; opening a door (Norman, 1988) is changing the door into a pass-through-able state. The empirical study of affordances typically emphasizes two questions: what precisely are the affordances that are used in the carrying out of a specific task, and how do we perceive them? The question of how we act on affordances is given less emphasis. Rather, the fact that affordances are clearly routinely acted upon by animals is what motivates the study of how these relations are perceived.

The affordances just mentioned, for step climbing, rock throwing, and door opening, are all examples of relations between individual people and individual objects. Clearly, many of the affordances that we act upon have a more complex structure. Some actions are actions controlled with reference to relations *between* objects. Placing a key in a lock is an example: what is being acted on is the relation between the two objects. In order to control this action the actor must attend to the relation between key and lock, coupling the action with the relation and monitoring as the distance between the two objects is reduced and finally eliminated as the key enters the lock. This kind of attending applies to the use of any kind of tool: using a shovel to dig a hole, using paint to change the colour of an object, or using a rock to crack nuts (as practiced by chimpanzees; Bril et al., 2009). And many of the actions we perform are already actions on relations between other people and external structure: spoon-feeding a baby, throwing a frisbee to another person, or pointing to an object. Notice that some of these actions involve physically manipulating an object (a spoon or a frisbee), others leave the object undisturbed (pointing). But all can be understood as actions controlled with reference to a relation between an object in the actor's environment and a second individual, also in the actor's environment.

3.3.2 Hearing: acting or being acted upon?

If speaking is acting on relations, what is hearing? Is it a type of perception, as is normally supposed, or is it a passive process of being acted upon, in which the addressee has no choice but to be manipulated by the actions of the speaker? In fact, neither of these statements is adequate to capture the status of the listener. Attending to another person's verbal actions is certainly a type of perceiving, but it is also actively engaging

with structured events and is therefore a type of action. And it is fair to say that a hearer can be acted upon by a speaker. If speaking is acting on relations then necessarily any given act of speaking must disturb at least one of the relata: the speaker must directly effect a change in either the object or the addressee, or both. But it is clearly not the case that addressees are merely passive non-actors in this process; addressees routinely refuse to comply with requests, disagree, feign ignorance, and so on. How can all of this be compatible with the description of speaking as acting on a relational web?

The key to appreciating the addressee's status is to emphasize that speaking is not acting on the addressee directly; it is acting on a relation which is connected to the addressee at one end. This means that while the speaker is acting, the addressee can simultaneously act on that very same relation. Refusing a request is then an act of vetoing the speaker's action. Presumably in a typical conversation a hearer will not bother to veto much of the speaker's verbal activity, but in a more hostile conversation the veto will be used more often. The addressee can also skillfully act to interpret the speaker's action for their own purposes, as happens when someone who is asked a question responds as if they have been asked a different question. And the conception of speakers' actions as actions on relations also provides a plausible means of accounting for what goes on in instances of misunderstanding. An addressee can perceive that a verbal action has been attempted, and can also see that it has failed (because their attention has not been directed anywhere in particular). This provides the addressee with a motivation to further explore their own relational web, or to act in turn to elicit further informative actions from the speaker (through the use of a clarification request, say).

Hearing, then, is an active process. It is a type of perception-action: a hearer acts in order to perceive structure in their environment, and perceives in order to act. And a cooperative addressee is not simply acted upon, but is allowing themselves to be acted upon, and actively participating in this process by attempting at each stage to attend to the appropriate parts of their own web. Speaking requires simultaneous engagement with structure in the world by at least two perceiving-acting individuals.

It would, however, be a mistake to assume that the relationship between speaking and hearing is the same as the relationship between acting and perceiving. An individual is a perceiver-actor, but not a 'speaker-hearer' (this phrase has nevertheless

appeared in the ecological literature; Hodges, 2009). In fact, speaking is itself a type of perception-action. And so is hearing. Speaking is acting on a relational field, but it is also attending to the consequence of that action. Hearing is actively attending to verbal structure: the hearer must actively attend to the actions being carried out on their own relational web if they are to be able to respond in an appropriate way.⁷

The fact that a speaker's verbal actions directly affect the behaviour of others echoes something that is already well appreciated within the study of animal communication. This can be seen in the assessment/management perspective, as proposed by Owings and Morton (1998). This combines two ideas: that animals produce signals in order to change the behaviour of other animals to their own advantage (management) and that at the same time they monitor their environment for opportunities to act and for threats from rivals and predators, and so on, looking out for the signals that are most likely to be reliable (assessment). Owings and Morton's innovation is to suggest that animals are able to *manage* the behaviour of other individuals precisely by exploiting the *assessment* activity of those animals. That is, assessment and management are two different kinds of activity, but they are codependent; one cannot function properly without the other. The assessment/management framework, like the present account of speaking, is presented as an alternative to the information-transmission or information-sharing model. Assessment is exploratory monitoring of the environment's affordances, whereas management is the set of actions directed at arranging those affordances for personal ends.

3.4 An outline of a theory of language learning

The proposal is that verbal actions are actions upon the relational field. Clearly, this raises a number of important questions. If these relations have to exist externally, how is it that we can talk about fictional objects, or tell stories? What are we acting upon when we produce a piece of writing, or when we talk to ourselves in our heads? How is it possible to speak to multiple people at the same time: does this not mean that

⁷Note that this is not a process of interpreting or of reconciling conventional meaning with context; the hearer is in effect seeking to answer the question: Which part of my relational web is being acted upon? Or equivalently, What is my attention being directed to? (Although of course the hearer is not *actually* asking themselves this question.)

the speaker must be performing a separate action on the individual relational field surrounding each addressee? If speaking is acting on an individual's relational web, why do all of our separate webs end up resembling one another, such that we all speak basically the same language as the people around us? How is it that we can produce metaphors, indirect speech acts, deliberately ambiguous statements, and deceptive utterances?

These are all good questions, and I believe they all admit of plausible answers, although I will not attempt to address them all here. The key to answering all of them will be to insist that skilled speaking is something that can *only* be understood with reference to the individual's development. One's ability to speak a particular language is *entirely dependent* on one's history ('speaking a language' is to be understood here in a narrow sense, as denoting the ability to produce verbal structures that are compatible with at least some of the relations surrounding other speakers, and excluding such things as pointing gestures and laughter).

Above I brought up the example of an anthropologist and an informant with no language in common. It was suggested that what is lacking in this encounter is the ability, on the part of either participant, to direct the other's attention to objects in the world in an efficient way. I now propose that what is lacking is the ability to act upon the other's relational web directly using verbal actions. When verbal actions are ineffective, individuals can still direct one another's attention to things, but in a more laborious way: by pointing, gesturing, pantomiming, and so on. And by doing these things, the individuals are able to learn the verbal actions that will work as effective attention-directing devices. That is, we are able to learn the 'names' of things by exploratory action with an engaged addressee. The name of an object is perceived by acting on or attending to a relation between another speaker and that object. It is important to note that the 'name' is not a property of the object. The object affords naming not because of anything intrinsic to the object but because of a correspondence between the name and a relation: 'naming an object' is really directing someone's attention to an object by acting on the relation between the object and the addressee.

The anthropologist-informant encounter represents one type of language learning: second language learning, carried out by mature adults who already possess the basic skills for speaking. Clearly learning of a more fundamental sort must happen much

earlier on, in infancy and early childhood. The pre-verbal infant finds itself situated in a richly structured field of relations and events. It must learn the meaning of this ecosystem through exploration. What must the infant first attend to, in the earliest stages, if it is ultimately to acquire the ability to speak?

An important characteristic of the new realist ecosystems view is that it posits that the environment is structured in a nested manner. There are objects, then there are relations between objects, then there are relations between relations, and so on. Logically, then, one would expect that the infant must learn the meaning of the kinds of structure encountered at the lower levels before moving on to explore the structure in the levels higher up. The earliest stage of infant exploration should therefore be characterized by attention to the structure of events and of objects (or surfaces). After the infant has attained a certain competency at this level, there will be a shift to the next level up, and the child will begin to attend to how certain actions *cause* certain other events. In other words, there will be a shift from attending to things, to attending to relations between things. One important class of events that the child must explore is the set of verbal actions, which can themselves cause further physical events, although the exact nature of these causal relationships can only be approached through continued exploration: a given utterance made by the infant may at one time elicit a certain response from a caregiver, while at another time a seemingly very similar utterance may elicit a very different response, or no response at all.

An important milestone in infant language learning is the shift from making sounds to ‘naming objects’ (this shift happens around the age of nine months and has thus been dubbed the ‘nine month revolution’ by Tomasello, 1999). On the present proposal, this may be better characterized as a shift from attending to event structure at the lowest level (to what it sounds like when different mouth movements are made) to attending to what verbal actions can be used to *do*. The proposal is that a child at the one-word stage is not simply naming objects, but is acting on, or exploring, the structure of the relational field. A child producing the word ‘apple’ is not deliberately sending a message to a caregiver (to be translated as ‘you give me the apple’, or some such), but is actively trying to bring about a change in its environment. The child may be attempting to bring the apple closer, or else may simply be attempting to disrupt the system in some way, to see what happens when the word ‘apple’ is produced. The message-

like nature of the utterance is really an artifact of the standard description scheme (to reiterate: the proposal is that speaking is *always* acting directly on relations; what is being claimed here is not that young infants lack an awareness that what they are doing is sending messages, but rather, that speaking is *never* sending messages, insofar as messages are objects sent through a conduit). If producing one-word utterances is acting on relations between objects and other people, and is not just producing the name of a thing for its own sake, then a prediction that follows is that children should produce more one-word utterances when there are other people around compared to when they are alone. Or at any rate the utterances that they do produce should be behaviourally distinguishable in the two situations.

In essence, the present proposal conceives of speaking as a kind of technique for pointing to things. At the one-word stage, the things being pointed to are objects. There should follow a developmental trajectory: first, the child points to objects, then to relations between objects, then to relations that are not immediately visible, then to entirely fictional relations that are created by the very act of pointing. Each transition can be conceived as a change in the state of the animal-environment system; the later states cannot simply be accessed from the beginning—the states must be passed through in sequence (learning is of course to be viewed as change in accordance with the doctrine of the new ball, not as the addition of symbols to a slate).⁸

The idea that relations can point to non-present or fictional objects is one that immediately demands further explanation. The suggestion, for now, will be that the solution to this problem lies in the developmental trajectory of the individual child. At the earliest stages, in order for a child to be able to learn a new ‘object name’ (i.e. a new verbal action), it will be necessary that both the object and the addressee are present at the point of learning. Later on, through exploratory use of the newly acquired action,

⁸In fact, there is some evidence in the cognitive developmental literature that learning does indeed follow something like the trajectory outlined. [Christie and Gentner \(2010\)](#) report results from a set of relational match-to-sample studies carried out with 3-year-olds. In these studies, children are shown picture sets AA (two tortoises) and BB (two cows), and asked to choose between two possible matches: CC (two rhinoceroses) and AB (one tortoise, one cow). The children in the study routinely picked AB as the matching pair (matching objects), unless they were given a name to refer to the higher-order relation (sameness), in which case 57% of children reportedly picked CC as the match (matching relation). [Gentner \(2010\)](#) uses these and other results to propose that children’s learning about relations involves a bootstrapping process: children learn terms for relations at a lower level of organization, and this gives them access to higher-level structures, which they can then refer to and reason about before moving up to still-higher levels.

the child will learn that the relation exists even when the object itself is no longer visible (e.g. uttering the word ‘ball’ will sometimes result in an adult leaving the room and returning with ball in hand). As the child’s verbal repertoire expands, the actions the child is able to carry out will be less and less confined to the immediate surroundings; the child will gradually learn that what is important for carrying out successful verbal actions is not that the referent object is immediately visible but that the addressee is present and is responding in an appropriate way to the verbal actions being produced. At this point, the child is already in a state where they should in principle be capable of referring to non-existent entities. In short, it is possible to talk about things that are not there only as a consequence of being able to talk about things that *are* there.⁹

There is an important distinction in all of this that has so far been left implicit in what has been said about learning: it is the distinction between learning to speak and learning to be spoken to. This distinction is a source of minor embarrassment for advocates of the cognitivist view of language as an internal system (Hendriks and Koster, 2010). There can only be one internal system of language, it is assumed. What an individual speaker can comprehend and what they can produce should therefore be identical, and if any discrepancy between the two is observed in reality, then this is a problem that requires an explanation. If there is precisely one system of language in each individual’s brain, then certain aspects of children’s verbal behaviour are puzzling. For instance, children over-generalize morphosyntactic inflections, effectively inventing words that no-one else around them has ever spoken (‘I goed’ instead of ‘I went’; Rumelhart and McClelland, 1986), and they show clear signs that they understand more words than they themselves produce (Benedict, 1979; Goldin-Meadow

⁹This description is consistent with Holt’s (1915)’s notion of the ‘recession of the stimulus’, as described in chapter 2. Vygotsky (1962) outlined a similar process. In particular, Vygotsky spoke of the child’s learning a language, and specifically learning to use private speech, as a process of ‘internalizing’ a behaviour first learned in interaction with others. Speech is said by Vygotsky to emerge first as public behaviour: the child learns to respond vocally to goings-on in its surroundings. Later, the child learns to use words as tools to control his own ongoing tasks. For example, the child might first learn to count by simply repeating a sequence of sounds. Later, continuing to say the numbers out loud, the child might use this sequence in another activity—perhaps while drawing, as a tool for keeping track of whether he has used all of his coloured pens yet. Later still, it becomes unnecessary for the child to actually vocalize the numbers. The speech is still *there*, but it has become ‘private’, or ‘inner’. It should be clear that the notion of ‘internalization’ here does not necessarily have to be understood in terms of the formation of a mental representation. What is internalized need not be some abstract symbol, but the concrete act of using a piece of structured behaviour to maintain control of an ongoing activity.

et al., 1976). Cognitivists' explanations for these behaviours appeal to such things as processing demands on language performance, or to competing systems of rules and representations (dual-route models and the like), or to social factors outside the scope of a mechanistic explanation.

For the theory currently being proposed, these phenomena do not present a problem. If it looks like there is a difference, in observed behavioural data, between language production and language comprehension, that is because there really is a difference. In Owings and Morton's (1998) terms, one is management, the other is assessment. Language is not conceived as a system in the head; rather 'language' is a label applied from the outside to describe a set of behaviours that take place in a populated environment. In terms of behaviour, acting on someone else's relational web really is a different kind of thing from having your own web acted upon. This is the case even though we can classify the actions that are used in acting and in being acted upon as being instances of the same words or structures. The difference is the same as that between watching someone else playing a violin with some level of skill and trying to play the violin oneself: it is not enough just to observe a set of actions; to become a competent language-user or violin-player one must practice those actions at first hand.

But if language production and language comprehension are two different skills, this might appear to leave unexplained why it is that we end up speaking roughly the same language as the people around us. Why do we not just make up our own language for acting upon other people's relational webs? The answer is fairly straightforward. We *could not* just invent our own language. None of our addressees would be able to respond to it, because they would lack the necessary reflexes and skills to be able to perceive what it is that their attention is being directed to. In reality, a learner will continue to produce the verbal actions that are effective while abandoning or refining those that do not seem to elicit a satisfactory response. In short, the reason we end up speaking the same language as the people around us is the one given by Pickering and Garrod (2004): language use in its most basic form occurs not in internal thought but in interaction with others. Or to put this in the terms introduced in chapter 1: the ecosystem in which the child finds itself is a niche that has been constructed through the activities of previous generations. And part of the structure of this system is constituted by the (adult) individuals themselves, who have become adapted to respond

to particular verbal actions. The child may in time alter the structure of this niche in some ways (it may invent new words or names for objects, which adults may learn to respond to), but the verbal niche is, in the most important sense, pre-constructed, and it is to this existing structure that the child must learn to adapt.

3.5 Putting the relational field model of language to work

There is much that remains to be worked out about the present theory if it is to provide a useful means of describing actual verbal behaviour. And of course there would be little point in attempting to re-conceptualize language as a conduit-free activity unless the new description can be put to some practical use. The broader aim of this thesis is to bring collaborative activities inside the purview of a sound nonrepresentational theory of psychology. A major advantage of the present proposal may be that it provides a means for making sense of the verbal actions that are used in collaborative situations. It provides a means of avoiding the problem that arises for conduit-based models of language: how do the language systems of individual speakers come to entrain with one another in practice, in a dynamic environment? On the present proposal there is no need to invoke an internal language system in the first place: utterances are conceived as verbal actions controlled with reference to relations in the speaker's environment, with the function of directing another person's attention in a useful way. Verbal actions are *part of* the activity, not parallel to it.

In order to assess whether this is a viable description scheme, it will be necessary to test it against empirical data. Part 2 aims to define an empirical methodology that will be suitable for analyzing collaborative activities in a manner consistent with the tenets of the ecological realist approach.

3.6 Postscript: Columbus crosses the ocean

I have been challenged to give an account of what happens when one hears a sentence such as 'Columbus crossed the ocean in 1492.' (This particular sentence was invoked

by Bertrand Russell in his *A History of Western Philosophy*, 1945, in the chapter on Dewey; Russell cites it as an example of a ‘true’ sentence.)

The first point to note is that this sentence does not mean anything except to someone who has already learned or become sensitive to the particular sound structures involved here. Sensitivity here means that the listener, upon hearing the given sound structure, experiences what the words are ‘about’, in some sense, this experience being invoked either automatically, by the immediate response of the listener’s nervous system, or with some effort on the part of the listener to discover the meaning of the structure being heard.

The second point is that to be able to make use of this sentence the listener must already have a good deal of social and historical knowledge. The listener must be familiar with metric time—the invented system for counting days, years, etc., within which ‘1492’ is known to refer to a year in the past. The listener must have previously learned the historical fact that America was discovered, by a man named Columbus, who sailed there from Europe across the Atlantic. Without these particular pieces of knowledge, the listener could only experience a vague sense of *somebody* crossing *something*; in this case, the listener might well conjure up on the spot a quite fanciful understanding of who ‘Columbus’ is and what he was doing (as children will when exposed to words that are not anchored to some existing concrete experience).

The third point is that the listener must have become sensitive to grammatical properties of the sentence, such as that the past tense inflexion is related to things that happened in the past, that crossing an ocean means sailing in a boat from one land mass to another, that the order of the words means that Columbus moved, while the ocean stayed where it was.

All of this is structure that is ‘out there’ and has to be discriminated, attuned to, by the learner. It is all very well saying that this is a true sentence and that it has grammatical structure, but we also have to acknowledge that interpreting this sentence relies on a great deal of learning. To understand this sentence, the hearer must have gone through a long Vyotskian or Holtean process of change—of internalization, or of causing the stimulus to recede.

The final point, then, is that this sentence is not a good model for understanding what language is or what speaking is in its most basic form. We cannot hope to achieve

a complete analysis of *how* this sentence is understood at the individual level, because to do so we would have to know an awful lot about the history of that individual. This might be seen as a severe problem were the goal of an ecological science of speaking to be to produce a complete, internally coherent account that could be used, say, to programme a robot to speak. But such is not the goal here. The goal is to be able to understand real-world practical problems in terms that can allow useful solutions. The question is not, ‘How can we build a robot that can understand this sentence?’, but, for example, ‘How can we help this child, who is having difficulty understanding this sentence, to be able to grasp its meaning?’ It is hoped that the focus on basic interactive processes will serve, as with the dynamic assessment approach to second-language teaching (e.g., [Poehner and Lantolf, 2005](#)), as a useful guide for identifying the particular source of the problem, and to encourage the child’s development in the appropriate direction.

Part II

Methodology: studying activity in a populated environment

Chapter 4

The task-oriented approach in psychology

Jerry Fodor's essay *The Modularity of Mind* (1983) was an attempt to address a major problem in cognitive psychology: if the aim of the research programme is to come up with an account of how the brain works, then the problem space is impossibly large: where does one start?¹ What is needed is some means of breaking down the overall problem into manageable pieces. Fodor believed that the only plausible means we had of dividing the problem space in a useful way was to endorse some version of faculty psychology: the old idea that the brain is by nature divided into special-purpose units, which on Fodor's proposal are modules—encapsulated computational devices that interact with one another but otherwise behave autonomously. If the brain is organized into modules this is good news because it means we do not have to study the entire brain, we can start by studying individual modules in isolation: 'The condition for successful science... is that nature should have joints to carve it at: relatively simple subsystems which can be artificially isolated and which behave, in isolation, in something like the way that they behave *in situ*. Modules satisfy this condition...' Fodor argued that the best candidates for modular processes are input-analysing systems such as vision and language, which were held to function inferentially, turning noisy and impoverished sense data into useful data in a format which can then be fed into central

¹ A version of this chapter is published in the *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (Baggs, 2014).

cognitive processes such as thought and memory (which are not themselves modular in nature).

The ecological realist, by contrast, is not attempting to give an account of how the brain works. The brain is not the object of inquiry. Rather the ecological realist project aims to give an account of an animal-environment system: the whole system comprising the animal in its environment. But a similar problem arises: where to begin? Again what is needed is a way of breaking down the empirical project into tractable research problems. The solution that ecological psychologists have settled upon is not to break down the system into component pieces, but to break down the things the system does into individual units. The solution is to study individual *tasks*.

This raises the question: what is a task? It will be argued that there are two quite different ways of answering this question, which correspond to the two types of psychology just identified. The cognitivist, or brain-oriented psychologist is quite familiar with the task as a research device. These tasks are usually carefully constructed laboratory-based activities which seek to measure some behavioural response to a set of stimulus materials. The task is devised as a means of indirectly measuring some aspect of the brain.

The ecological or task-oriented psychologist, by contrast, aims to study real phenomena: activities that people actually engage in outside the laboratory. A prerequisite for being able to study a task in this way is that the task itself can be defined in a precise manner. The phenomena that have been studied most extensively by ecologically-inclined psychologists are simple action-control phenomena: how does a driver control steering, how do infants learn to walk, how does a baseball fielder run to the correct position to intercept a ball in flight? All of these might be considered as tasks in some sense, although it is not equally clear in each case where we should place the boundaries separating the task from other background activity. The baseball outfielder problem is neat in that it has clear start and end points, and a clear criterion for successful performance: the catcher consistently ends up in the right place to intercept the ball. But where does an individual steering task end? And when is the task of learning to walk complete, if ever?

The present chapter has two aims: 1) to achieve precision about the concept of a task, around which some ambiguity exists; in particular I will pursue a definition

of the concept that will allow us to expand the scope of the ecological realist research programme beyond simple visually-guided phenomena like the outfielder problem, but without losing the rigour and precision that such activities impose; and 2) to argue that the task, when defined in an appropriately precise way, is a unit that satisfies the criteria for a solution to Fodor's problem: how can the overall research project be broken down into manageable pieces?

4.1 The task as an epistemological device

The first question we must address concerns what nature of thing a task is: which category should we place it in? Should the task be considered a part of reality, or a part of our description of reality? Is the task a thing that resides somewhere in the system, or is it a tool for describing that system in some way?

Within the rationalist framework, a case could be made for either response. Within the ecological realist framework however, only the latter response makes sense. This can be seen if we try to apply the concept of a task to a specific activity, say kicking a ball. On the traditional, internalist view, one might be tempted to say that the task is something that resides inside the actor. In order to be able to engage in the activity of kicking a ball, the actor must on some level be able to categorize what they are doing *as* an act of kicking a ball. And further, in order for a particular action to count as an act of kicking, the actor must initially *intend* to carry out that particular action. On this view, the task is perhaps construed as an internal plan or recipe which exists in the actor's head prior to the carrying out of the action: it is a part of reality, and not just of the description.

This rationalist manner of construing the concept of a task is subject to the same criticism as was made by Gilbert Ryle (1949) in relation to the concept of 'volitions'. If a task is an internal entity, how might one go about individuating separate instances? If tasks are real, discoverable things, then it should be possible to answer such questions as the following: 'What was the last task you completed?' 'Of the tasks you have fulfilled today, which took the longest?' 'How many tasks must be performed in preparing a French onion soup?' And then there is the problem of how a particular task or plan is selected. If the actor, in performing some task, must first select a particular

method for performing it, this is an internal act of choosing which appears to constitute a task in itself, a task which must itself have been selected somehow. The whole thing is thus threatened by an infinite regress. A possible escape, for the internalist, is to say that the actor does not choose the task, but merely acts out whatever script is currently presenting itself. But this implies that the actor must be an automaton—a perennially unpopular proposition.

Within the ecological realist framework, by contrast, there can be no recourse to positing internal entities. The system is the animal in its environment. In the ball-kicking example, the system consists minimally of an actor and a ball and a perception-action relation linking the two. There is nothing in this system that we can label as the task. Perhaps the relation itself is the task? This might work for a system with only two entities and one relation, but now suppose the actor is kicking the ball at a tree. There are now at least three relations in the system (linking actor and ball, actor and tree, and ball and tree), none of which we could point to individually and label as *the* task.²

We are left with the conclusion that the task does not name some component of the system but rather identifies a particular way of looking at the system from the outside. The task is an epistemological device. Some happy consequences follow from adopting this position. It means that we do not have to worry about how to correctly individuate particular instances of tasks: there is no correct way because there is not anything out there to individuate to begin with. Instead, we can define a task in whatever way is convenient for some purpose.

But as we saw above with the examples of catching a ball and learning to walk, there is much diversity in the things that can potentially be labelled as tasks. The problem is how to define the concept of a task in a useful way, in a way that can be applied quite generally. And in a way which allows us to draw useful conclusions from the particular phenomenon under investigation.

²I here assume that three-way relations are not possible, or else must be reducible to a set of binary relations. I take it that if we allow for the existence of higher-order relations then it becomes possible to define a task at an arbitrary level of complexity (to describe the same activity in an arbitrarily large number of ways), and it again becomes impossible to identify a single thing that constitutes *the* task.

4.2 The acting animal and the task-seeking psychologist

First, though, it is necessary to add a qualification to the assertion just made. If the task is defined as an epistemological construct, then it is a thing that is strictly artificial, invented for the benefit of the researcher. However, it is not the case that a researcher is justified in labelling just anything as a task. Nature must still possess joints to carve it at. In order for the analysis to be valid, the researcher's description of the task must correspond to what the actor is in fact doing. The task is a third-person device, but it must recognize that the actor occupies a particular first-person perspective.

Failing to take into account the first-person perspective of an actor is an old problem which William James named the psychologist's fallacy. The fallacy is committed when a researcher assumes the validity of a particular description of what the subject is doing, even though the subject may have a very different view of things.

This is illustrated by Martin Orne's 1962 work on research subjects' behaviour in hypnosis sessions. A question that arises for hypnotists is: when patients behave as if they are hypnotized, how does one know if they are really hypnotized? Might they not be perfectly aware but only simulating hypnosis for some reason (to avoid embarrassment, for example)? To investigate this, Orne tried to trick his subjects: he attempted to create the impression that the hypnotist had left the room during the session and that the subject was no longer being observed. A number of the subjects were convinced by this act and gave up the pretense that they had ever been hypnotized: they began moving about or even got up and wandered around the room. By doing so, they demonstrated clearly to an outside observer that they had in fact been simulating all along. What this demonstrates is that it is not sufficient for a psychologist to merely describe a situation as it appears from the outside: a third party observer may be inclined to take at face value any behaviour that looks like behaviour under hypnosis. Only by considering what the situation looks like from the point of view of the research subject is it possible to assess whether the third-party description is a valid one.

In what follows I have attempted to reserve the term *task* for referring to the researcher-defined unit of study, and to talk of the actor as acting or carrying out actions.

4.3 The task as a reconfiguration of resources

The discussion so far has attempted to identify what a task is not: it is not an entity in an actor's head, nor is it an arbitrary description drawn directly from a psychologist's intuitions. We must now attempt a definition of what a task is.

The solution that ecological psychologists arrive at is to define a task as a particular, characteristic reconfiguration of resources. That is, the task should have a precisely defined start and end point, and a recognizable mode of transition between the two. This is important because it provides a means to unite different instances of a single phenomenon as being instances of the same task, and not just a disparate set of individual things that happened. The word *resources* here can be defined quite broadly; the purpose is to pick out any component of the animal-environment system that is used as an instrument in completing the task, or that changes in a characteristic way over the course of the activity. This might be a change in some object or other structure in the environment, or might be a change in the animal itself. Thus, a ball-catching task is any instance of a catcher attempting to locomote from some initial position to a position where they can catch a ball which is following a parabolic trajectory through the air. A steering task is any instance in which a driver must control the motion of a car around a bend. And the task of learning to walk is characterized by a change in the infant, from being incapable of ambulating between one standing position and another to being capable of doing so.

The claim is that if we can characterize a specific reconfiguration that happens somewhere in the system, then we have defined a task. Subsequently, whenever the relevant conditions obtain, we are justified in characterizing what is happening as an instance of that task.

Notice that these task descriptions also specify criteria for successful completion of the task. The task is completed successfully if the ball is caught, if the car negotiates the bend without disaster, and if the infant eventually learns to walk. Furthermore, the fact that the task is defined as a precise reconfiguration of resources means that we also have a way of assessing whether we can further apply the analysis of the phenomenon under investigation to some novel phenomenon that has not itself been studied. Thus, the analysis of ball-catching should apply for catching any projectile that follows a

parabolic trajectory through the air, but not for catching a frisbee, and not necessarily for how goalkeepers in soccer move to stop bending free kicks.

It might be unclear whether learning can really be characterized as a task in the same way as the other activities. In the case of learning to walk, it looks like what is happening is a local change within the infant. One may be tempted to suggest that there is no reconfiguration of physical resources here—that the infant already has all of the necessary resources to be able to walk, and that what is needed is really a change in mental structure, which cannot be characterized from the outside. This would be quite wrong, however. It would be wrong partly for the obvious reason that any change in mental structure must also be a change in physical structure somewhere. But more importantly, learning to walk is not a process that occurs only inside the infant, but is one that involves the entire animal-environment system. The infant never exits the environment to practice in private. The way infants learn to walk is by exhaustive trial and error: repeatedly trying a few steps before falling down and trying again ([Adolph et al., 2012](#)). The falling down is part of the learning process, and the only way to tell whether one has fallen down is by attending to information from the environment. If what is being reconfigured is the animal-environment system and not just the animal, then the task of learning can in this instance be characterized from the outside after all.

Perhaps there are some activities that people engage in that cannot be characterized from the outside, and thus must remain out of reach for a psychology committed to empirical investigation based on the present definition of the task. An operation carried out entirely ‘in the head’ may be one such case. An example would be someone working through some particular sum without then making the solution ‘external’ in any way. Here the initial sum is being reconfigured into a solution, but there appears to be no means by which a psychologist could measure the outcome, and thus no way of characterizing this from the outside as a task. This is certainly a limitation, but then again it’s not clear why a psychologist would want to study this kind of thing: why bother asking people to work on sums if you are not going to check the answer in some way? Such operations carried out ‘in the head’ must remain elusive for empirical purposes, but this need not necessarily be a source of great concern.

More likely is that we can identify some phenomenon informally that we would like to define as a task, but which turns out to be of such complexity that it is unclear

where to posit a start and end point, or how to characterize what goes on in between. This presents an important practical limit on current empirical study, but also motivates the project of developing better tools for characterizing a broader range of tasks.

Wilson and Golonka (2013) outline a four-point procedure for applying the task-oriented approach to the general case: 1) identify the task: conduct a task analysis ‘which characterizes from a first person perspective the specific task that a perceiving-acting cognitive agent is faced with’; 2) identify task-relevant resources available to the agent which can potentially be assembled in the carrying out of the task; 3) ‘identify how the agent can assemble these resources into a system capable of solving the problem at hand’; and 4) recruit research subjects and have them perform the activity in order to test whether they do in fact carry it out in the fashion identified in (3). Clearly step (1) is crucial: it is necessary to define the task in a precise way at the outset if steps (2)-(4) are to be possible. We are presently concerned with what is presupposed by the first step: what does it take for something to count as a task? The suggestion is that the things we can characterize as tasks, and therefore the things we can study, are the things where we can identify a characteristic reconfiguration of resources with a clearly identifiable start and end point. If this formulation of the task is a good one, we can hope to be able to use it to extend the scope of the task-oriented approach beyond phenomena of the ball-catching type to encompass more complex activities.

4.4 Acting with other actors

A particularly acute test of the methodology will be how well it can be applied to activities involving more than one actor. This presents a notorious set of difficulties for psychologists working in the brain-oriented tradition: if action is the output of mental processes, how can it be that some actions appear to be carried out by multiple actors? Cognitivists have attempted to resolve this by appealing to various mechanisms, including internal prediction processes, mirror mechanisms that are held to transmit mental content by contagion, and group minds (Sebanz et al., 2006).

The task-oriented approach does not encounter this difficulty, because it does not posit that behaviour is the result of operations in a distinct mental realm. But there remains some ambiguity to be resolved. If a task is a third-person device for describing

a first-person activity, what then are we to make of actions that are carried out by multiple individuals? If two people are rowing a boat together, how many tasks are there? Well, since the task on the present definition is a descriptive device rather than an aspect of reality to be discovered, the question is ill-formed: there is not any pre-existing number of tasks to be untangled. The number of tasks there are depends on how we choose to define what constitutes a task.

Might it be useful to define a task as a thing carried out by the pair: the pair is the actor, and the task is what this pair does? This does not seem plausible. An individual is an experiencing perceiver-actor, an organism that goes about perceiving invariant structure in order to control action. A group of people is not an experiencing unit, and does not appear to possess any means of controlling its own actions, except through the actions of the individuals that make up the group.

In the rowing example, then, there must be two tasks. That is, we can define what is going on as two individuals each carrying out their own task. For one rower the task might be to set the pace and to carry out equally-spaced strokes with their oar, while for the other rower the task may be to match the first stroke-for-stroke, applying equal force on the opposite side of the boat. In this example, the pair will either succeed or not depending on whether or not the two rowers are able to coordinate their respective activities. But the apparent success of the pair is really a consequence of individual success on the part of the second rower. Of course, in reality things may not be so simple: the first rower may actively coordinate with the second, instead of merely setting the pace and letting the second do all the work of maintaining coordination. But the point is that as long as the individual tasks are defined in an appropriate way, there is no need to appeal to any additional process at the group level. The group level activity should already be defined in the individual tasks.

And to reiterate, the individual tasks are not private processes going on in the heads of the actors, but processes that span the animal-environment system. The distinguishing feature of a task that is carried out with others is that the environment happens to contain other animals who are also producing structure that is relevant to the task at hand.

4.5 Rationale for the task-oriented approach

Psychology has in recent history been defined as the science of mind, brain, and behaviour, or some combination of those three things. In any case, the unit of analysis is routinely taken to be the individual organism. Tasks have thus been understood as things that psychologists get subjects to do merely as a proximate source of data about the real phenomena hidden in the head. What reason could there be to give up on this formulation and to instead carve up the problem space into individual tasks, making the task the main unit of analysis?

An important rationale for the task-oriented approach comes from work on learning. Learning in infancy is shown to be thoroughly context-dependent, the child's general abilities arising out of immediate experience with local problems (Thelen and Smith, 1994). This context-dependency applies across a wide array of different activities, and at different points of development. One example comes from how infants learn to negotiate slopes when they are crawling versus when they have begun to walk. Crawling infants initially attempt to descend a slope head-first, even if the slope is too steep to negotiate. Through further exploration these children gradually learn to adapt their approach to slopes and will become increasingly cautious when faced with a steep descent. This cautiousness does not, however, transfer straightaway to the new task of walking: infants who have newly begun to walk are once again unable to perceive that a slope does not afford walking down and will attempt to descend when encouraged by an experimenter. This happens even though the same infants are perfectly capable of perceiving that the slope does not afford descent if they are placed in a crawling position at the top of the same slope (Adolph et al., 1993). In a sense, the transition from crawling to walking means that the infant has to learn what its environment means all over again.

Thelen and Smith (1994) cite further examples of context-specific learning, including work on the 'shape bias' in children's object-naming, which is a phenomenon that emerges when children are specifically asked to name objects but not when they are merely asked to match objects with 'similar' targets; that is, the shape bias is dependent on the context of the experimenter's demands (Smith et al., 1996). A similar context-dependent process is observed in normal language learning, in which children appear

to initially use any given verb in only a handful of specific syntactic structures—the so-called ‘verb island’. Only later do they begin to generalize, using novel verbs in syntactic structures where they have not previously been used (Tomasello, 2000).

And this task-specificity in learning does not appear to be unique to young children. Evidence for similar context-specificity in later life learning comes from studies of brain training products which claim to promote improvement in general cognitive abilities. These studies have shown that although users of these products improve at the specific tasks that they have to perform as part of the ‘training’, there is no evidence that this learning transfers to novel tasks (Owen et al., 2010).

So learning is context-specific, and thus one rationale for doing psychology one task at a time is that this does in fact appear to correspond to how we learn things. But does this not suggest that tasks are, after all, genuine components of the system, and not merely analyst-defined units? Tasks are the things we learn to do? No. This is just another confusion of description and reality. Contexts, like tasks, are not naturally distinct from one another. It is only when we look in from the outside that we can divide things up in this way. From the perspective of the learner, there is no thing which constitutes the context of the present, there is only the present, which is continuous with everything else the learner does. The point is only that the task-oriented description scheme may better correspond to reality than the brain-oriented scheme.

A further rationale comes from empirical success. The task-oriented approach has in fact been shown to produce hypotheses that better correspond to reality than those generated by the alternative general problem-solving approach. To return to the ball-catching example, the task-oriented approach here treats this as a problem to be solved by the catcher in a simple way: the catcher maintains visual attention upon the ball in flight in such a way that the ball appears to follow a linear path up and down (McBeath et al., 1995). This predicts that the catcher will run in a curved line to intercept the ball, in contrast to the general problem solving approach which predicts that the catcher will perform an internal calculation in order to work out the optimal place to run to and will run to that spot in a straight line. In fact, catchers do follow a curved path before catching the ball, lending empirical support to the task-oriented hypothesis and indicating that the internal-calculation hypothesis is false (Wilson and Golonka, 2013).

A final rationale here is a rationale from usefulness. Applied fields exist to solve

practical problems. Take speech and language therapy as an example. This field is not organized around the question of how speech is organized in the brain but around what can be done in practice to improve patients' speech in a measurable way. Treating children with articulation disorders is, for the speech therapist, a two-part task consisting of firstly improving the child's articulation of specific sounds, and secondly improving the child's systematic use of those sounds so that the child's improved pronunciation can be generalized to novel words and not just to the words they have already practiced (Gierut, 1998). This particular treatment technique uses precisely a task-oriented approach, on the present definition. Treating an articulation problem as a task to be solved is useful, and even a necessary pre-requisite for devising an effective treatment methodology.

4.6 Final definition: what is a task?

The discussion in the preceding sections allows us to state the following as a definition of a task.

A task is an analyst-defined unit that corresponds to a recognizable reconfiguration of resources in an animal-environment system, which:

1. is meaningful from a first person perspective; it presents opportunities to act and constrains possible solutions,
2. can be defined precisely, has a start and end point, and a transition between them, that can all be directly observed,
3. specifies criteria for successful completion,
4. specifies conditions under which conclusions can be generalized to similar tasks.

I will here address a couple of possible objections to the task-oriented approach. Firstly, if the objective is to study individual tasks one at a time, is this not just massive modularity by another name? No. Massive modularity, like Fodor's non-massive version, still appeals to modules in the brain that are not directly observable. The task-oriented approach by contrast, is concerned with disruptions of physical resources that can in principle be measured directly: the physical resources *are* the task; they are not measured as a proxy for investigating a hypothetical construct in the brain.

Secondly, if the task-oriented approach is all about coming up with useful characterizations of specific, circumscribed phenomena, that is all well and good, but it can never lead to a complete picture, either of how the brain works or of how the animal-environment system works. I think this objection is perfectly valid, except that it characterizes as a limitation what could equally be seen as a strength. The strength is that the task-oriented approach leads to immediately useful conclusions. By contrast, suppose we had a perfect model of the brain, assuming such a thing to be possible. That model would be a monumental scientific achievement, but it would still only be useful for solving particular questions we might ask it—that is, for addressing specific tasks. The task-oriented approach is a means of not having to wait for that model to be finished, it is a tool for solving practical problems in the present. In fact, this is a similar conclusion to the one [Fodor \(1983\)](#) arrived at. Modularity, it was hoped, might work as a methodological tool for studying peripheral input systems such as language and vision, but it cannot give an account of central cognition or thinking: ‘The ghost has been chased further back into the machine, but it has not been exorcised’.

I will make one final point. A current debate in cognitive science opposes a traditional symbol-manipulating disembodied view of cognition with various purportedly embodied alternatives ([Wilson, 2002](#); [Shapiro, 2011](#)). Or more exotically, the view of the mind as an internal property of an organism is opposed to the ‘extended mind’ which is said to span brain, body, and environment ([Menary, 2010](#)). These debates pit two apparently incompatible sets of ontological claims against one another: the claim that minds are properties of individuals versus the claim that minds cross the boundaries of skull and skin. This has led to fairly fruitless discussions on the question of *what* we should call ‘cognitive’: can external symbols constitute mental content or must mental content be confined to the head? In the present chapter I have emphasized a different kind of dichotomy, on methodological rather than ontological lines, between a brain-oriented and a task-oriented approach to doing empirical psychology. I suggest that this difference in methodology is real and unambiguous, and that the dichotomy is a useful one. Making the dichotomy on methodological lines allows us to concentrate not on what our ultimate model should look like, but on *how* to make empirical progress towards it by investigating real phenomena that we are interested in.

The case in favour of the task-oriented approach ultimately rests on its usefulness. The approach may serve as a useful general framework for psychological study, if the concept of a task can be defined in a precise way. The present discussion suggests that it can be, and further proposes that the task is a viable solution to Fodor's problem of how psychological research can be organized. The challenge that arises is the following. If we want to study psychology in a tractable way, we have a choice between two options: there are modules, and there are tasks.

Chapter 5

A relational field–based account of referential communication games

Psycholinguistic research on dialogue recruits a standard empirical device: the referential communication game.¹ These games take a variety of different forms, but the structure is basically the same across all instances (Yule, 1997). What happens is that an experimenter invites a pair of research volunteers into a laboratory and gives them a problem to solve together. We will call these volunteer subjects S1 and S2 (in practice one of these is often an experimental confederate). Typically, S1 is given a set of experimental materials and S2 is given a somewhat different set of materials, and there will be a physical barrier separating S1 and S2 such that each can see their own materials but not those given to their partner. The problem that they are asked to solve requires that at least one member of the pair somehow communicates something about the structure of their own materials to their partner who does not have direct visual access to that structure.

The format of these games was initially developed for addressing a set of research questions about feedback between speakers and addressees in communication (Krauss and Weinheimer, 1966). The earliest studies were explicitly carried out within the framework of information theory (or communication theory), which posits that language use should be understood by analogy with how signals are sent between two

¹These are also commonly called ‘referential communication *tasks*’; however, since I have already assigned a specific technical meaning to the term *task*, I will avoid re-using it here. I will refer to all of the laboratory-based activities discussed in this chapter as ‘games’.

telephones. The conduit-based model of speaking was presupposed, and the format of the referential communication game was built upon this presupposition.² The game format seems to be an attempt to reconcile two things: 1) all communication (by hypothesis) is sending messages back and forth through a conduit, but 2) communication in a normal environment is noisy and full of feedback loops—and it is therefore difficult to characterize what is going on from the outside. Since what was desired was a means of studying communication empirically it was thought necessary that the complexity of normal communication should be reduced, or controlled for. The solution was to create a simplified communicative set-up, one that maintains the conduit while removing the feedback loops that were not deemed by the researcher to be of interest. Thus, in the game set-up, access to shared visual structure is denied and the only things that may be exchanged between research subjects are words (and potentially other gestures). But what this means, in effect, is that these early researchers justified their model of communication as a telephone by creating a game format which can itself be played over a telephone line. In short, the format that these games take is strongly influenced by the theory that guided their development.

Because of this, I do not believe that referential communication games are a suitable tool for developing the relational field model of communication in a practical direction. The problem is that these games are artificial. They are artificial in at least two ways. Firstly, they impose artificial constraints on the actions available to S1 and S2. This is not such a problem; such constraints are necessary in a great deal of empirical enquiry. The problem is the second type of artificiality: the problems that subjects are expected to solve—the *task* that an individual is faced with—does not appear to correspond to anything outside the laboratory. The discussion of the task in the previous chapter asserted that the aim of our empirical research should be to identify

²Although the game setup has remained broadly constant over the years, the interpretation of the participants' behaviour has evolved considerably. Krauss and Fussell (1996, p. 661) provide a useful discussion of this. They note that, 'The Encoder/Decoder model, and at least the indirect influence of information theory, can be seen in the terminology of early studies of verbal communication.' They maintain that while this has had a lasting influence on the rhetoric used by human communication researchers, it has not particularly constrained how researchers interpret their own results (Krauss and Fussell, 1996, p. 661): 'information theory has not contributed importantly to the study of human communication. The aspect of the theory that has had greatest scientific impact is its ability to characterize information in an abstract, quantitative way. And a major impediment in using the theory to describe human communication is that, for a particular message transmitted at a particular time to a particular receiver, more often than not we are at a loss to specify just what uncertainty (if any) has been reduced.'

precise mechanisms by which particular tasks are completed. Conclusions drawn from research on some particular task can then be generalized to other situations where the same mechanism is employed. In the case of referential communication games, the task to be completed by an individual does not appear to resemble anything much that is encountered in day-to-day speaking. As such, the conclusions that can be drawn from studying these games may not be generalizable to anything except other referential communication games (and perhaps certain conversations held over the phone).

So why discuss these games here at all? The answer is that these games have been studied quite extensively and have produced a fairly rich set of data. We can take this data and try to make sense of it within the framework of the field model of speaking. The point is not to apply the framework to the data for its own sake, but to use the data as a means of exploring what the framework can do, and what it will be useful for. In particular this chapter will have two aims. Firstly, I will again try to drive a wedge between two different approaches: the conduit-model of speaking and the relational field model outlined in chapter 3. The first part of this chapter will demonstrate how these two approaches lead to two entirely different analyses of the general structure of referential communication games, and therefore two genuinely different ways of thinking about what is going on when we speak. Secondly, I will look at some of the best-established findings drawn from the study of these games and I will attempt to re-analyze these findings within the relational field framework. The aim will be to test the limits of what the framework can be used for. Although there may be no practical upshot from analyzing these artificial games, nevertheless, these games are widely employed by researchers studying dialogue and interaction. For the field model to prove its worth, it must be able to provide an analysis of these games. Moreover, since it is generally taken for granted that these games can *only* be analyzed within a conduit-based description scheme, it will count as a major achievement for the relational field model if it is able to provide a plausible re-analysis that invokes no conduit at all.

5.1 The structure of referential communication games

The field model proposes that speaking should be understood as a kind of action that is inherently bound up with local circumstances. There is nothing novel or controversial about this; indeed, to point out that speaking must be understood in context is to state an almost perfect platitude. Herb Clark (1996) identifies a broad school of thought within language research which he calls the ‘language-as-action’ tradition. Roughly, this is the set of research programmes that are built around precisely the idea that language use should be understood in context. This is set up in contrast to the ‘language-as-product’ tradition, which attempts to understand language as a formal system in its own right, which Clark identifies with work by Chomsky and followers.

Historically, however, research in the language-as-action tradition has routinely assumed a conduit model of information exchange. The desire to understand speaking in context inevitably comes into conflict with the use of the conduit model as a descriptive tool. The conflict is that the conduit itself does not appear to be context-specific. Contexts can change, but the conduit itself and the content that passes through it must always exist in a separate, de-contextualized form, to be interpreted at either end with reality as perceived by the current addressee. As a consequence, researchers working in this tradition face a dilemma. There arise two distinct types of research programme that can be pursued. The first insists on a firm commitment to the importance of context, and thus tries to describe in great detail what is going on in a handful of very specific scenarios. This can produce useful descriptions of the particular scenario under study, although such descriptions have not tended to provide generalizable insights about the mechanisms involved. The second programme, by contrast, quickly loses sight of context; the goal instead is to explain the nature of the conduit itself, and the nature of how speakers use the conduit. It is in pursuit of this goal that researchers have developed the referential communication game.

That the second programme is quick to forget about context is best seen in the way it constructs its theories of communication. These theories are often accompanied by some variant of the diagram in Fig. 5.1 (or else, the diagram is implicitly assumed). In the diagram, there are minimally two speakers and a conduit. This is presented as a generalized abstraction that can later be applied to any given context. The theory is

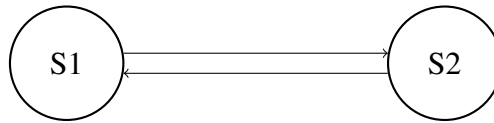
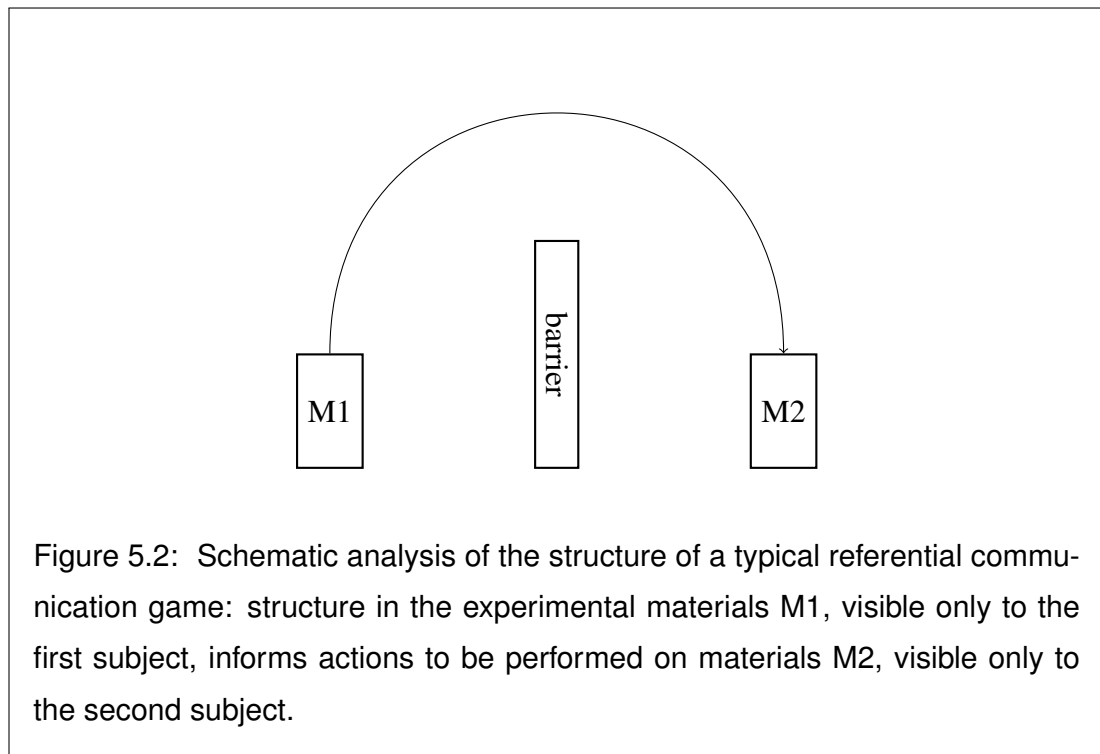


Figure 5.1: The conduit model of speaking: speakers transmit information to one another encoded in sound waves.

then built around this basic diagram. The specific theories that result appear to express a quite diverse range of concerns. [Clark \(1996\)](#) devises a ‘ladder’ of collaborative actions which he says allows speakers to infer the present content of mutual knowledge; [Tomasello et al. \(2005\)](#) emphasize how mutual knowledge is actively managed by speakers through the initiation of episodes of joint attention; [Pickering and Garrod \(2004\)](#) posit that each speaker is in possession of an automatic processing module that keeps track of the current situation on the speaker’s behalf; [Sperber and Wilson \(1995\)](#) place the burden of adjusting for context upon the speaker, who is said to take the present circumstances into account before producing an utterance. What all of these have in common is that they assume that the difficult part of the problem is in *how* precisely the conduit is used. The reasoning appears to be that, since a) there is only one conduit, and since b) this conduit is present in all of the contexts that we’d want to generalize to, therefore c) an adequate theory of language-use will follow if only we can devise an appropriate theory of the conduit. In effect, the proposals mentioned above are all theories of conduit-use.

5.1.1 The extended conduit model

It is instructive to try to apply the diagram in Fig. 5.1 to the set-up of the referential communication game in its general form. We will start with a task analysis of a basic game. Some structure on one side of the barrier, visible only to S1, has to be duplicated on the other side of the barrier. S2 must reconstruct the target structure using a set of disassembled materials. This general format is depicted in Fig. 5.2. In order to solve this overall problem, both S1 and S2 must engage with structure in the environment



somehow. The question is, how?

Taking the conduit model and superimposing it upon a bird's eye view of the game set-up, we get the diagram in Fig. 5.3(a): S1 and S2 are connected to one another via a communicative conduit, and each has a perspective view over a particular subset of the experimental materials, M1 and M2. The arrows pointing from S1 to M1 and from S2 to M2 represent the subjects' attending to the external materials. Fig. 5.3(a) thus represents the state of affairs just after S1 and S2 have entered the laboratory (and after they have been given whatever instructions the experimenter gives). Now the subjects must set about trying to solve the task. Typically, the subject who has been assigned the director role, S1 say, will start by consulting their materials, constructing a verbal instruction and uttering it out loud. This verbal instruction is depicted in the Fig. 5.3(a) as the arrow pointing from S1 to S2—this is one of the two arrows that make up the speech conduit. Next, S2 interprets this instruction and attempts to reconcile it with an action they can perform upon M2. Once S2 is satisfied that this is the correct action, they will carry this action out, effecting a change in the structure of M2. This action is depicted in Fig. 5.3(a) in the arrow pointing from S2 to M2.

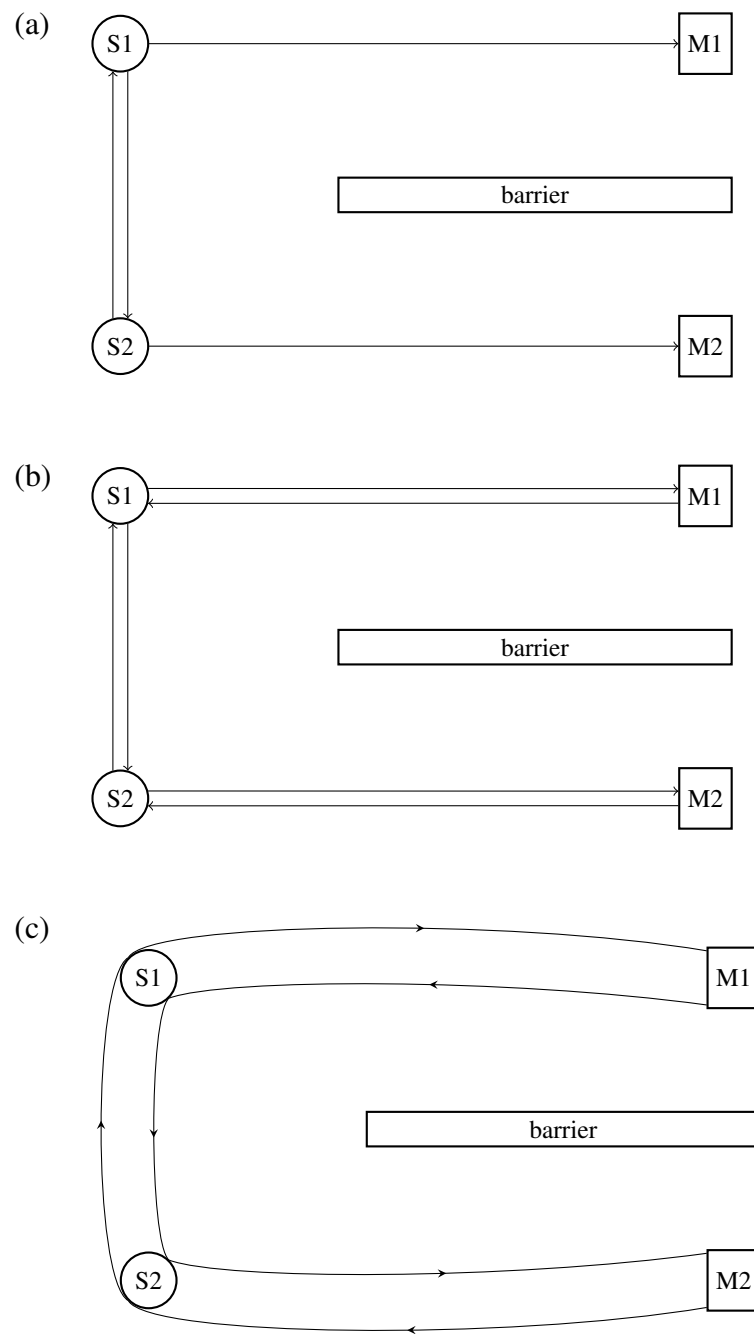


Figure 5.3: The game understood according to the conduit model: (a) the initial set-up consists of a truncated conduit (S1–S2) plus structure in the world; (b) feedback creates an extended conduit; (c) information flow through the extended conduit

But now suppose that S2 has attempted to reconcile the instruction from S1 with structure in M2 and has been unable to select a suitable action. In effect, S2 has probed the structure in M2, and this act of probing has created or revealed information for S2 to the effect that no suitable action is available (we will assume that the information here is Shannon information). This is represented in the arrow in Fig. 5.3(b) pointing from M2 to S2. S2 needs more information. S2 will therefore send a message back down the conduit to S1, requesting further instructions. Next, S1 must consult M1 again: attending to the structure, collecting further information about that structure (arrow M1 to S1 in Fig. 5.3(b)). S1 is now equipped to produce a refined instruction, which will again be sent down the conduit. The interaction continues in this cycling fashion until the subjects are satisfied that they have reached a point of termination.

Fig. 5.3(b) is a static diagram which appears to have a fairly simple structure. There are just six arrows connecting everything together. Looking at Fig. 5.3(b), it is tempting to think that any one of these arrows can easily be pulled out of the whole and examined in isolation. However, the diagram is only useful if it can capture how things work in motion. These games, after all, are activities that exist in time as well as space. The description in the last two paragraphs of the perfectly routine way in which these games are completed suggests that Fig. 5.3(b) is misleading: there are not really six distinct arrows that can be isolated from one another. Instead, the whole system works as a kind of extended conduit or circuit. Information can be seen flowing from M1 to S1, then to S2, M2, then back again as an ambiguity arises and further information is requested. The information flowing through any one arrow at any one time is contingent on the structure of the entire circuit.

A less misleading depiction of the structure of these tasks is shown in Fig. 5.3(c): there are no distinct arrows; it is recognized that the entire system forms a continuous circuit around which information flows. But notice that this raises a problem if the aim is to use the referential communication game format as a means of studying the nature of the speech conduit, or the nature of interaction, or the nature of the cognitive system of S1, or of any other individual component of the circuit. Typically, an experimenter will put a set of materials into a laboratory environment, and will then invite in S1 and S2 and have them solve a problem with those materials. Some form of behavioural measurement will be taken, and the resulting measurements will later be

used as a source of data for addressing some research question about some component of the system. But if the game format allows continuous flow of information through a distributed system, as in 3(c), then the experimenter may not in fact be justified in attributing any particular aspect of the measured behaviour to a particular component in the set-up. Any given action or utterance made by S1 may be a behavioural response to information that originated within the cognitive system of S1, or it may equally be a response to information that ultimately originated in M1, or in S2's cognitive system, or anywhere else in the circuit. Because every component in the set-up is connected in one continuous informational loop, it is not possible, in practice, to identify any single component within this system as the definitive cause of a given measurement outcome.³

All of this is important because it relates to the issue of representative design: to what extent are conclusions drawn from the study of these games generalizable outside the laboratory? The fact that the extended conduit is causally opaque is not in itself a reason to reject the model. The problem is that it is not known to what extent the behaviour observed in these games corresponds to behaviour in the kinds of natural, everyday dialogues that are supposedly being modelled. The diagram in Fig. 5.3(c) aims to illustrate clearly that the materials that are placed in the laboratory by the experimenter must be playing a causal role in the behaviour that is ultimately observed. But the nature of the causal role played by these materials is unknown. Is the game set-up really a simplification of what happens in all dialogue, or is it rather an invented scenario that produces its own distinctive patterns of behaviour? This is a problem that is ubiquitous in psychological methodology; it was identified notably by Egon Brunswik (1956) in his work on vision. One way to address it is to insist, as Brunswik did, that the design of psychological experiments should be 'representative'; that is, investigators should sample over not only different items and different individuals, but also different situations (Hammond and Stewart, 2001b; Dhimi et al., 2004).⁴ Thus,

³This critique of the format of referential communication games parallels John Dewey's (1896) critique of the 'reflex arc concept' in psychological research.

⁴Historical note: the term 'representative design' was introduced by Brunswik in opposition to conventional or factorial design, which had initially been developed for use in agriculture rather than psychology. Subsequent writers have adopted the concept of representative design, but have given it the name that Brunswik coined for a different concept, 'ecological validity' (e.g., Orne, 1969; Neisser, 1976). For Brunswik, ecological validity was a property of the cues available in the proximal stimulus

results from referential communication games would only be relevant to a psychology of dialogue if their conclusions are additionally shown to generalize to other situations with different task characteristics. The ecological realist solution, as outlined in chapter 4, is to make narrowly-defined tasks the object of inquiry instead of broadly defined phenomena such as ‘dialogue’. It is at any rate an open question exactly how what goes on in a referential communication game relates to what goes on in dialogue in general. The conduit-based research programme has partly addressed this issue: the use of communication games as a research tool is supplemented by analyses of more naturalistic speech corpora, for instance (Schober, 2006). However, the issue is one that is yet to be fully addressed.

The fact that information must flow continuously around an extended system that includes subjects and materials is obvious when the game is depicted as in Fig. 5.3(a–c). Yet the referential communication game format is widely used without mention of this problem. Why is it this issue appears to have gone unaddressed? I think the answer lies in Fig. 5.3(a). I have drawn the arrows pointing outwards from the subjects to the materials, to represent a process of active attending. Actually, a more traditional conception of these games implicitly places the arrows pointing in the opposite direction, from material to subject, to represent the materials’ role as sensory *input* to S1 and S2’s cognitive systems. On this input-based scheme, all of the discussion that takes place in the carrying out of the game is local to the subsystem consisting of the speech conduit and of the two cognitive systems belonging to S1 and S2, each of which is held to contain a mental model of the situation. We will call this the truncated conduit model. In effect, the game set-up is understood to consist of a subsystem where language occurs, and a physical environment where actions occur. There are only two types of interaction between the language subsystem and the physical environment: one-way input of sensory data from environment to cognitive systems, and one-way output of behaviour from cognitive systems to environment. This notably overlooks the possibility that individuals do not maintain a complete mental copy of the entire scene as perceived, i.e. the possibility that the environment is used as a store of external representations to be consulted as necessary. This input-based scheme is committed to a complete dichotomy between language and environment. I suspect that this way of conceiving of

(light) for judging the presence of a distal object (see Hammond and Stewart, 2001a).

how the game works is inevitable if the two-heads model depicted in Fig. 5.1 is taken as a true depiction of how communication works.⁵

5.1.2 The field model

Now, how would an account of this game look if it were based on the relational field model? First of all, there are no arrows pointing in only one direction, from an object to a perceiver, say. The arrows in the relational field point in both directions: from a visually exploring perceiver-actor to an object, and simultaneously, via light, from the object to the perceiver-actor. These two-way arrows are relations. So the first thing to observe is that, at the beginning of the game session, S1 and S2 stand in a perception-action relation to M1 and M2, respectively. These perception-action relations are depicted in Fig. 5.4(a) by two bidirectional arrows, R1 and R2. S1 has direct perceptual access to M1 and S2 has direct perceptual access to M2, but of course the Ss do not have direct perceptual access to the M on the other side of the barrier. However, note that in the ecological realist framework, relations themselves are held to be directly perceivable. What this means is that each of the subjects also stands in a perception-action relation to their partner's perception-action activity; that is, S1 can directly perceive R2, and S2 can directly perceive R1. And further, according to the proposed model, each subject can also *act upon* their partner's perception-action relation. Notice, also, that there is no arrow pointing between S1 and S2 directly, as there is in the conduit model in Fig. 5.3(a–c). Of course, S1 and S2 can see each other, but the arrows are left out of the diagrams in Fig. 5.4, because they are not relevant. What is relevant is not that S1 can see S2 per se, but that S1 can see what S2 is doing *relative to* some other structure.

Since there are two subjects in this game, and therefore two first-person perspectives, there must also be at least two different tasks. At this point it will be useful to

⁵Perhaps there does exist some way to eliminate this dualism whilst maintaining the conduit. It is conceivable that a different description scheme could be built on the principles of distributed cognition, appealing to the concept of external representations. Input-output psychology recognizes only internal representations. The distributed scheme is essentially what is depicted in Fig. 5.3(a–c). In order to turn this distributed description scheme into an empirical research programme, the description would have to be accompanied by a practical methodology that could draw useful conclusions about the functioning of the extended system. How such a methodology could be developed is a problem that need not concern us here.

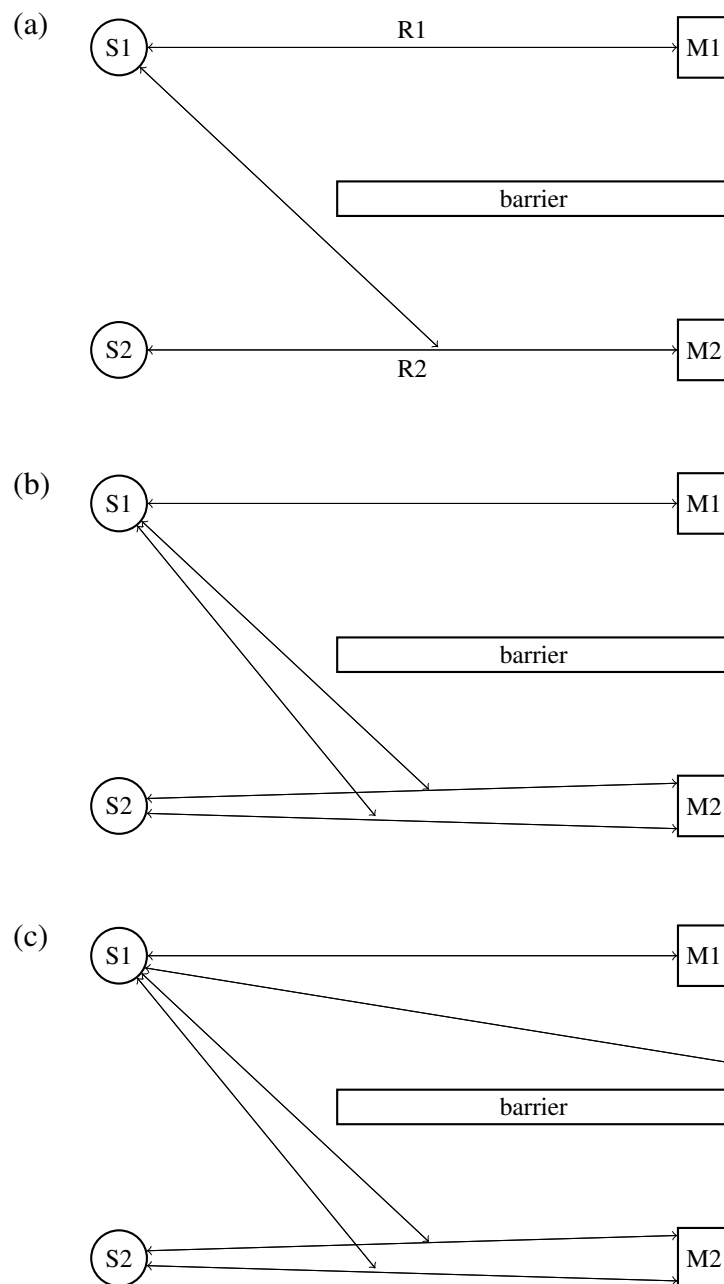


Figure 5.4: The game understood according to the relational field model, showing incremental learning by S1: (a) S1 gives the first instruction; (b) by attending to S2's behaviour, S1 comes to discriminate structure in S2's relational web; (c) this also generates information about the structure of the materials. (A similar diagram could be produced for S2.)

pause to examine what the task is for each subject. The following is a first approximation. For S1 the task is to use the structure in M1 as a resource for altering the structure in M2; to do this, S1 must use M1 to assemble instructions, and must direct these instructions at S2's relational web. For S2 the task is to use the materials that constitute M2 in order to assemble something that matches M1; in order to do this S2 must gather information about M1 via S1's relational web. Both of these tasks will terminate at the same time, when both S1 and S2 are satisfied that the structure in M2 has been made to match the target structure in M1 (this is the end of the task regardless of whether or not the structures actually do match).

The rest of this section should be read as a tentative effort to take what is so far a sketchy outline of a relational-field based account of speaking and to apply it to a particular type of verbal exchange, filling in some of the details along the way. Many of these details may have to be revised later.

Fig. 5.4(a) can be read as S1 (the instruction-giver) producing the first instruction in the game, which is to be understood as S1 acting on R2. Next, S2 must respond to this action. But how? I suggest that it will be useful to think of S1's instruction as an event that can have two possible outcomes. First, S2's relational web may already be appropriately *calibrated* to respond to S1's instruction; that is, S2 has already attended to certain structure in the environment and is now in a state of readiness to respond to certain actions that refer to that structure. In this case, S1's action *directly orients* S2 towards some aspect of M2. To say that S2 is directly oriented to this structure is to say that S2's attention is directed there by necessity; it is as though S1 had picked up S2's hand and placed it on the target object—S2 has no choice but to attend to that object. So the suggestion is that the instruction, by itself, orients S2's attention. But the instruction does not, by itself, put S2's limbs into motion in response. What it does is it effects a change in S2. S2 is turned into a particular kind of structure-seeking device. For S2, the first instruction is perceived as an event that is in some way informative about a change to be effected in M2. This event may be sufficiently informative that S2 is confident that they have enough information to act. Or else, S2's search for relevant structure will fail, at which point S2 may act in turn on R1, seeking further information.

The second possible outcome of S1's initial instruction is that S2's relational web will turn out not be appropriately calibrated. In this case, the instruction—that is, the

action on R1—will fail to direct S2’s attention to anything in particular. S2 is not turned into a structure-seeking device in the same way. S2 perceives that an action has occurred, but does not perceive the meaning of that action. Instead, S2 can only respond by seeking further information. There are many ways S2 can do this: pausing and waiting for further instructions, stating that the instruction is insufficiently informative, or asking a question in return. All of these might be understood as instances of S2 acting on R1 (including pausing: failing to respond may itself be a kind of action).

Let us suppose that S2 apprehends the first instruction, and tries to carry out an appropriate action in response. This attempt is itself informative, for S1, about the nature of R2. If S2 looks at *something*, this act of looking tells S1 that the particular ‘something’ referred to is in fact there in the environment (in M2). It also constitutes information for S1 about S2’s *awareness* of that something.

But the instructions in these games are not just about individual objects. They are about how to assemble a set of objects into a larger whole. This means that it is not adequate to picture R2 as a straight line connecting S2 and M2. M2 is not one thing, but a set of things. R2 is really a web. It is the web of relations that connects S2 to everything in the environment which S2 is currently aware of, or attuned to. S2’s behaviour in response to S1’s instructions is informative not just about one relation, but about a set of relations. S2’s behaviour is informative about the structure of S2’s web. This is depicted in simplified form in Fig. 5.4(b).⁶

Now notice that R2 does not exist in isolation from the rest of the environment. R2 is really a sub-component of a larger relational web which includes R2 and R1 and includes all of the other relations that abound between entities. This means that any disruption in the structure of R2 may potential be informative about other, connected parts of the web. In particular, it must be the case that S2’s behaviour is informative about the relationship between M1 and M2. S1’s instruction is itself an attempt to alter this relationship, and to bring about a relation of sameness between the two set of materials. S1’s perceiving of this M1–M2 relation is depicted in Fig. 5.4(c). In a

⁶The static diagram in Fig. 5.4(b) appears to suggest that S1 is acting on two relations at once. Is this necessary? I think not. Perhaps in practice it is better to stipulate that S1 can act on two relations in sequence, or else can on a higher-order relation which itself connects two more basic relations. For example, in the instruction to ‘Put the book on the chair’, S2 is directed to a relation between a book and a chair, and what S1 is acting on is a relation connecting *that* relation to S2.

sense, the M1–M2 relation is where the action is in this game. The task, for each of the Ss, appears, ultimately, to be about altering this relation. There is some work to be done to clarify how this can be the case.

Fig. 5.4(c) depicts S1 in a state of attunement to a set of relations in the environment. Does this not require short term memory and therefore internal representation on S1's part? It requires at least a temporary change in S1's dispositions. Presumably this involves some restructuring at the neural level. But this need not necessarily mean that something is being stored, like symbols on a slate. All that is entailed is that a change has occurred. For now, we will assume that this change is consistent with the doctrine of the new ball: S1's learning, or attuning to parts of the environment, is a temporary change in how S1 is disposed to act towards the environment.

What has been described so far, and what is depicted in Fig. 5.4(a–c), could apply to a referential communication game in which S1 produces instructions and can see S2, but S2 is not allowed to ask questions in return. Commonly, however, S2 will be allowed to freely produce verbal actions of their own. S2's actions are actions on S1's relational web, which are themselves a potential source of information to both S2 and S1. What this suggests is that there are a great many sources of potential information for both subjects in these games. It is not the case that the Ss can only learn about their partner's materials by extracting meaning encoded in the things that are said, as the (extended) conduit model supposes. Nor is it the case that linguistic information is a special kind of information that must be reconciled with a physical context through a process of inference. When the whole speaker-environment system is considered in dynamic terms, it is clear that the information that is available is richly structured and that it is informative about the environment broadly and not just about the partner's perspective on that environment.

It must be acknowledged that the field model is no closer to being able to give a general account of how referential communication games work than is the extended conduit model. In both cases, the system is complex: dynamic processes mean that any given component in the system causally interacts with all of the other components. As a result, the underlying causal structure is opaque. The advantage, for the field model, is that it escapes the problem of representative design: what causal role is being played by the structure of the experimental materials? The field model does not

suffer this problem because the experimental materials are part of the phenomenon to be explained. The materials are not merely tools recruited by a researcher as a means of investigating a broader phenomenon ('dialogue', or 'collaboration'), but objects of empirical study in their own right. The aim of the task-oriented psychologist is precisely to answer the question: how are *these* materials being assembled in the carrying out of *this particular task*?

5.1.3 A note on roles and repertoires

I have been referring to the volunteer subjects in these games by the labels S1 and S2. It is more common, however, to refer to the subjects by the names of roles that have been assigned to them by the experimenter. S1 is referred to as the 'instructor' or 'instruction-giver', the 'director', or simply the 'speaker'; S2 is the 'follower', 'matcher', or 'addressee'. I have avoided using these labels because they are, I think, theory-laden in a problematic way. To call S1 the 'director' is to imply that S1 has unquestioningly assumed a particular job-description for themselves. This obscures the exploratory, learning behaviour that subjects must engage in whenever they arrive in a laboratory for an experimental session: subjects do not simply arrive with an understanding of the game as it is understood by the experimenter; they must work out for themselves, to some extent, what it is that they are expected to do.

Perhaps it is merely pedantic to avoid using these role labels? It might be argued that these labels are just a convenient shorthand that denotes roughly what the task is for each subject. Still, a convenient shorthand can cause a lot of trouble. The temptation is to slide from using the role label as a description applied from the outside to talking about the role as part of the actual, causal machinery that brings the game to completion. It is tempting to assume that in order to produce instructions, i.e., to behave like a director, S1 must actually *be* a director, in some psychologically meaningful sense. But there is little reason to assume that people generally see themselves as performing some well-defined role (except in institutional settings, such as being the Speaker in the House of Commons, or being a priest presiding over a wedding ceremony). And in any case, the role labels that are typically used in dialogue games are rarely justified empirically: experimenters simply assign the role labels, without mak-

ing any attempt to verify or demonstrate that the role really does describe the subject's own understanding of what it is that they are expected to do. As a consequence, the literature on referential communication games generally contains little information about demand characteristics—about how tasks look from the subject's point of view. One means of assessing these task demands is through quasi-experimental control methods, such as post-session questionnaires; I have also proposed that audio recordings and session transcripts can themselves be used for this purpose (Baggs, 2013).

A more parsimonious alternative to invoking roles as causal mechanisms is to say that people build up a *repertoire* of actions and dispositions (Rogoff et al., 2006). On the present scheme, an animal's repertoire might be understood as the set of affordance relations an animal has established between itself and its environment.

5.2 Accommodating established findings

The literature on referential communication games is characterized by an entrenched set of disagreements about how, generally, to go about interpreting the data. There are broadly two camps: mentalizers (e.g. Clark, 1996) and automatizers (e.g., Pickering and Garrod, 2004). For present purposes, it will not be useful to get into the details of the disagreement. Mentalizers and automatizers all agree that speaking should be understood as a phenomenon involving mental content. The disagreement is over how that content gets there and how it is used. These are questions which do not make sense for the relational field model. In this section, I propose to examine a set of findings that are said to be 'robust' across different games studied by different researchers; I will examine four 'basic findings' identified by Schober and Brennan (2003), plus a fifth which I have added (clarification requests; Purver et al., 2003).⁷

1. With repetition, referring expressions become shorter and more definite, and articulation is reduced.
2. Speaking in dialogue is different from speaking in monologue.

⁷Actually, Schober and Brennan identify five phenomena themselves, but I have collapsed their items (1) and (2) into the present item (1).

3. Speakers appear to design at least some of their utterances for the needs of a specific audience.
4. Overhearers to a referential communication game are at a disadvantage relative to the two participants proper.
5. Speakers produce clarification requests.

I will outline how each of these phenomena can be understood as activities in the relational field.

5.2.1 The reduction of referring expressions

The first robust finding is that, where Ss in a game must refer to the same objects repeatedly, the length of the referring expression tends to decrease. Thus, in [Clark and Wilkes-Gibbs's \(1986\)](#) game, which used a set of tangrams (peculiarly-shaped shadow figures) printed on cards, S1 on the first turn might refer to 'the one that looks like a monk praying', then on subsequent turns the expression is reduced until it eventually becomes simply 'the monk'.

This phenomenon can perhaps be quite easily incorporated into the relational scheme. Presumably, it does not make sense to define a task whose first-person aim is to produce a referring expression per se. Rather, these expressions are only recruited as instruments in a larger task (that is, a first-person activity directed at bringing about a particular change in the environment). As such, there's no reason to continue producing the long referring expression—that is not the point. The point is to get the other person's cards in the right order (or to get to the end of the session as quickly as possible, or whatever).

But why, precisely, do the expressions become shorter? Perhaps it is more pertinent to ask why it is that the earlier expressions are so *long*. This question is one that can, I think, be answered. In producing the first instruction, S1 is acting on R2, but this action is partly exploratory. The materials that the subjects have been presented with are novel, and these items have been deliberately selected as things that do not already have names. What this means is that there is at the beginning no efficient, one- or two-word action that S1 can perform on R2 that will unambiguously direct S2's attention

to the appropriate card. The best that S1 can do is to direct S2's attention to some feature of the target figure that disambiguates it from the other figures. S1 thus has to direct S2's attention to structure in M2 in a particular way. In the example above, S1 is highlighting a resemblance between a particular card and 'a monk praying'. Notice that this also affords S2 an opportunity to learn: S2 will now be disposed to perceive the card as a monk as well (even though S2 may not have noted the resemblance prior to the action from S1). This means that on the next occasion when the subjects need to refer to this card, their dispositions have been shaped by the previous action. Assuming that the Ss' disposition to see this card (and no other) as a monk has not decayed in the intervening time, there is now no ambiguity in referring to 'the monk'. An affordance has been created for efficiently referring to a unique item.⁸

A similar phenomenon is observed at the level of articulation of individual words: the first utterance of a given word will tend to be longer and more clearly pronounced than subsequent instances (Fowler and Housum, 1987). Also, S1's articulation is generally shorter when S2 is present than when S1 is asked to speak into a tape recorder for the benefit of an absent S2 (McAllister et al., 1994). The explanation for this shortening of articulation must follow roughly the same lines as the one for the shortening of syntactically-structured referring expressions above. In the normal version of the game, subjects are probably not trying to produce a series of perfectly articulated patterns in sound waves for the benefit of the experimenter's tape recorder, rather they are trying to get a task completed.

But there is a difference between the explanation at the articulation level and at the syntactic level. It is not the case that an individual *word* has to be created. At the articulation level, there is no need to create a new affordance from scratch. When S1 produces the sound pattern 'monk' or 'ice skater', S2's relational web already contains structure that can be acted on by these sound patterns—the affordance is already there. However, there's probably nothing in the immediate situation that the term 'ice skater' points to specifically. That is, the relational web (R2, and also R1) has to be calibrated

⁸Incidentally, this explanation also makes sense of the finding that, when S1 receives no feedback about S2's behaviour, the referring expressions do not become shorter with repetition (Krauss and Weinheimer, 1966; Hupet and Chantraine, 1992; Garrod et al., 2007). The reason for this is that, under these conditions, S1 has no means of perceiving that an affordance for efficient attention-directing has been created. In the standard version of the game, S1 cannot perceive changes in M2. In the no-feedback condition, S1 cannot perceive changes in R2 either.

to the local environment.⁹ Perhaps there need be no particular psychological-level for the reduction in length of articulation. The initial articulation of ‘ice skater’ is embedded in a larger process of calibration. It may be that the initial utterance is longer as an unintended consequence of this higher-level calibration process. The first instance of ‘ice skater’ is an action that is performing two jobs: it is directing S2 to something in M2, but it is also calibrating R2 relative to a particular piece of action structure: the sound pattern. For R2 to be calibrated, S2 must engage in a process of effortful learning: S2 must discriminate the particular object that is being referred to as ‘the ice skater’ from all of the other possible candidates. Perhaps there are several figures that could plausibly be the ice skater, and additional information is initially required to pick out a single figure. This also means that S1’s initial instructions must be assembled with some care. In an example included in [Clark and Wilkes-Gibbs’s \(1986\)](#) report, the first utterance in fact does provide additional (syntactic) structure that is presumably helpful for guiding S2’s discrimination activity: ‘All right, the next one looks like a person who’s ice skating, except they’re sticking two arms out in front’. Prior to this initial utterance, R2 has not been calibrated, and thus S2 is unlikely to be able to respond efficiently to the term ‘ice skater’. After the first utterance, things are more efficient: subsequent utterances do not need to do the double work of calibrating and referring at the same time. The first instance of ‘the ice skater’ is an act of *creating* a name; subsequent instances are acts of *using* a name. And the first requires more effort because there is a greater degree of ambiguity when referring for the first time. The suggested explanation for reduction at the *phonological* level is that the utterance is embedded in a larger, effortful activity of name-creation, which is another way to say that a relational web is being calibrated. Admittedly the phonological-level explanation is quite speculative, but it points towards a potentially useful way of thinking about the multiple-level dynamics involved here.

⁹Another way to phrase this would be to say that the verbal action has to be ‘grounded’ in the local circumstances. Of course there’s a difference between this kind of grounding and grounding in [Clark and Brennan’s \(1991\)](#) sense. Grounding here is attaching structure in the relational web to something in the environment. Grounding for [Clark and Brennan](#) is building up a shared store of mental content.

5.2.2 The difficulty of monologue

One item in [Schober and Brennan's](#) (2003) list of fundamental findings is the claim that speaking in monologue differs from speaking in dialogue. This appears to be motivated, for these authors, by some results already mentioned above, showing that referring expressions (in a referential communication game) do not become reduced in the same way if a partner is present compared to if no partner is present. But the claim that monologue is different is really a weaker version of a more interesting claim: that monologue is *difficult* because it is a special case of dialogue that requires additional skills ([Pickering and Garrod, 2004](#)). 'Monologue' here denotes the production of linguistic structure intended for an audience that is not currently present, or that cannot interject. It includes such things as delivering a sermon, writing a letter, and leaving a message on an answering machine. It is readily apparent that all of these activities require special practice: an ability to converse fluently with others does not entail an ability to write coherently or to deliver an engaging lecture. Can the relational field model account for this?

Perhaps the main reason why monologue is difficult is that it is a form of acting in which the perception-action link is broken. Speaking is acting on relations, but in monologue it is not possible to perceive the consequences of these actions: perception is delayed or denied. Or rather, a particular kind of perception is denied: perception of the behaviour of others in response to the action. The speaker in monologue must compensate for this partial decoupling of perception and action by employing a different kind of strategy.

An analogy can be made here with carrying out an individual activity under conditions of perturbation. For example, suppose you wanted to compose a hand-written letter in the dark. Here, again, is an instance of action partially decoupled from perception. The task is different from writing under normal conditions, in a well-lit environment. In the dark, the writer still receives haptic feedback from the pen making contact with the page, but visual control of movement is not possible. As it happens, it is surprisingly easy to write a line of text without looking. The difficulty is keeping the line straight, and then moving on to a new line and starting at the right place. The new line is especially tricky. One way to consistently move down to a new line suc-

cessfully is to mark the start of the previous line with one's non-writing hand. This finger-marking technique is a special skill that can be employed to compensate for the lack of visual feedback. The suggestion is that for the speaker in monologue such special skills must be called upon—skills that are not routinely employed in more typical dialogue circumstances.

What precisely are the equivalents of the finger-marking technique for speaking in monologue? Clearly there are a great many skills to be learned; this is what makes monologue difficult. Specific devices that are used include formulaic or rote verbal structure—think about the kinds of messages people leave on answering machines. External aids such as scripts or notes are used by sermonizers, but rarely by people in dialogue. Writers must pay close attention to the structure that connects sentences together. In particular, new structure has to be anchored as it is introduced in the writing. In dialogue, there is no particular constraint preventing a speaker from abruptly changing the topic of conversation: the addressee will either follow or will be confused and will respond with a clarification request (see below); in this way, any difficulty is quickly resolved. A writer does not have this fallback, and so must carefully structure the verbal actions (the words and sentences) so that the reader will be able to follow. An ecological description of writing might go like this: writing is plotting a journey through the relational field; written words are stored actions; for these actions to be effective, the writing must be structured so that each step in the journey is a move from one anchored piece of structure to another. How it is that we are able to do all of this is clearly a matter open to substantial investigation.

5.2.3 'Audience design'

A source of some rancour within the literature is the question of the extent to which speakers take into account the needs of their conversational partner when performing an utterance (Brennan and Hanna, 2009). It is claimed that at least some speech is partner-specific, and that any such partner-specificity is evidence of 'audience design'. I put this in scare quotes because I think the way that this is commonly phrased begs an important question: it assumes that a speaker, in designing a particular verbal action for some particular individual, must take into account the *perspective* of that individual.

Of course, this is another manifestation of the conduit model: if a piece of linguistic structure is something that an addressee must reconcile with a context, then the speaker must design the linguistic structure in such a way that it is sufficiently informative about how this reconciling should be carried out. ‘Audience design’ implies design for a listener’s *understanding*. An alternative account might acknowledge that utterances are designed, but claim that what the speaker is concerned with is the effect that the verbal action will have on the *behaviour* of the person addressed—the person whose web is being acted upon.¹⁰

On the conduit model, it looks like in order to produce an utterance designed for a specific addressee, the speaker must perform some mental operation to work out how the current situation appears from where the addressee is standing. For the field model, this internal operation is not necessary. Speaking is acting on a relational field, and relational fields are inherently shaped by the presence of addressees. In fact, according to the field model, speaking in dialogue is *always* specific to an addressee.¹¹ In the account of language learning outlined in chapter 3, it was suggested that the earliest utterances are actions on specific relations between specific addressees and specific aspects of external structure. But the idea that all speech is partner-specific is obscured by the fact that our relational fields come to resemble those of other speakers of our language: for a large number of the verbal actions we make, there will exist many individuals whose relational webs are calibrated so as to be compatible with that action.

Experimental investigation of audience design uses a variant of the referential communication game set-up in which S1’s conversational partner is switched part-way through the session (Brennan and Clark, 1996). The standard finding is that S1 will continue to use a verbal action that has been established or anchored with S2, e.g. some object will be consistently referred to as ‘the shiny cylinder’, but S1 will switch smoothly to a different referring expression with S3, assuming S3 speaks first and produces a different expression, e.g. ‘the silver pipe’ (Metzing and Brennan, 2003). We

¹⁰For one such approach, see Thompson’s (1997) natural design perspective. Thompson conceives of design as an association between two arrays—an array of structures (verbal actions, say, ‘Can you pass the bread’) and an array of uses (outcomes that the speaker is trying to achieve—to acquire bread). In the case of communication, the speaker is said to be attempting to cause the addressee to enact a further design (to cause the addressee to enact the action of bread passing).

¹¹This is not the case in monologue, however, which suggests another reason why monologue is difficult: in monologue we must attempt to restrict ourselves to actions that are *not* partner-specific.

will assume that these results can be taken at face value. The field model has a natural explanation for the observed phenomenon. What speakers are acting upon is a relation between S1 and S2. The hypothesis is that S1 can perceive that S2's web is calibrated to respond to the first expression, and can equally perceive that S3's web is not appropriately calibrated to respond to it. The information for perceiving these two facts is manifested in the events that S1 observes: appropriate responding from S2, and the use of a novel action by S3. (Whether S1 in fact *makes use of* this particular higher-order structure is a matter that must itself be established empirically.)

Is this behaviour not precisely an instance of 'audience design'? That is one way to describe it. But actually things are more complicated than that: the game set-up itself is encouraging S1 to attend to higher-order structure in the relational field (to structure that specifies that S2 is in a state of relation with referring expression X, and that S3 isn't). Clearly whether S1 will bother to attend to this kind of structure must depend to some extent on the precise details of how the game is set up, and of how S1 apprehends what it is that they are being asked to do. These are details that are not typically included in published write-ups (Baggs, 2013).

Partner-specific speaking is a phenomenon that is encountered outside the laboratory too. An interesting case concerns bilingual speakers who come into contact with monolingual interlocutors.¹² Bilingual children appear to be sensitive to the abilities of their monolingual addressees from as early as age 3 (Romaine, 1995). The field model hypothesis says that using different actions with different addressees is just using different actions to act on different relations. If a bilingual speaker does not code-switch with a monolingual addressee, this need not involve any kind of deliberate perspective-taking. Rather, the speaker's actions are a practical response to the way the field is presently perceived: the web surrounding the monolingual addressee is perceived as affording a certain set of actions and not others.

5.2.4 The overhearer's disadvantage

Schober and Brennan's (2003) final fundamental finding arises from a study by Schober

¹²Curiously, little research attention appears to have been given to this, despite the fact that code-switching in bilinguals was one of the motivating concerns in the paper that originally introduced the term 'audience design' (Bell, 1984).

and Clark (1989). This was a version of the tangram card-sorting game (Clark and Wilkes-Gibbs, 1986), except in addition to S1 and S2, there was a third participant, S3, who had the same set of cards to sort as S2 but was not allowed to interact with either of the other subjects. The finding is that S2 consistently performs better than S3, even though S3 hears the same exchange. Schober and Clark's explanation for this is that, whereas S2 can respond to instructions that they find ambiguous by requesting further clarification, S3 cannot. When an instruction is ambiguous to S3, and no further instruction is forthcoming, all S3 can do is guess which card is being referred to. Schober and Clark conclude that 'understanding is part of a collaborative process', and that S3's disadvantage is that they are not really part of the interaction.

A slightly different explanation can be constructed using the field model. We might think of the overhearer's disadvantage as arising not from the fact that S3 is not part of the interaction per se, but from the fact that they cannot actively explore the relational field and are thus not able to perceive sufficient structure to complete the task. Perception-action is a first-person phenomenon; it is a tenet of ecological realism that one must in general act in order to perceive. An overhearer cannot act in the same way: they cannot act in order to generate further information for themselves about something that is temporarily ambiguous. In fact, the overhearer's disadvantage is the inverse of the problem of the difficulty of monologue: monologue is action partially decoupled from perception; overhearing is perception partially decoupled from action.

But there is a second reason for the overhearer's disadvantage: S1 can monitor S2's perception-action activity, and can potentially detect miscalibration of S2's relational web. Thus, S1 can act to correct mistakes made by S2, but not to correct mistakes made by S3.

There is some a priori reason to prefer the field-based explanation here. The major point in its favour is that it is framed entirely in terms of the perception-action activity of individuals. S2's advantage arises from two sources: S2's control of their own exploration activity, and S1's ability to act to correct miscalibration in S2's web. (In principle it should be possible to separate these two processes empirically.) Schober and Clark's conclusion, by contrast, appears to rest on the idea that there is something special about collaboration itself. They appeal to the idea that S1 and S2 are constantly monitoring one another's understanding. But this kind of claim invariably raises the

spectre of infinite regress: how do I know that you've successfully monitored *my* understanding? The field model does not require constant other-monitoring. What is monitored is whether actions lead to successful outcomes. When a mistake is detected, S1 and S2 can engage in further exploratory action, through which their relational webs are recalibrated. It is this calibration activity that is doing the explanatory work. The difference between S2 and S3 is that miscalibration in S2's web can be corrected by further action by S1 or S2, while miscalibration in S3's web can only be corrected by S3 on the basis of information already available via S3's perceptual systems.

5.2.5 Clarification requests

The final item I will discuss here is not included in [Schober and Brennan's \(2003\)](#) list, but I include it here because it poses an interesting challenge for the field model. When we are unable to respond to someone's verbal action, we often act in turn seeking clarification about what that person meant. These clarification requests take various forms: some are formulaic actions that can be recognized as clarification requests even when lifted from the supporting context ('what did you say?'), others appear to depend more on the immediately preceding utterance ('Can I have some toast please?' / 'Some?' / 'Toast'; see [Purver et al., 2003](#)).

The problem, for the field model, is that these clarification-seeking actions appear to reach back in time: they appear to refer to an immediately-preceding action which has now passed. If the field model aims to provide an anti-representational account of speaking, this is something it must account for. If verbal actions are actions on relations, what are clarification requests? Are they actions upon actions? How can this be? A representational account of speaking can appeal to structure stored in active memory: what a clarification request refers to is this temporarily stored structure. The field model, it would appear, cannot.

But perhaps this is a false problem. Perhaps it is created by the description of time as something that comes in discrete slices, like frames in a motion picture. This is a description of time that has been inherited from Newtonian physics. It has already been claimed, in [chapter 1](#), that perceiving at the psychological level is a temporally extended activity. And if perception can be temporally extended, then action must

extend in time as well: perception and action are, after all, two ways of looking at a single phenomenon.

But then again, this is exactly what is meant by ‘working memory’: working memory is a device for enabling perception and action to extend across time. There is a deeper issue here which ecological psychologists have yet to properly address: how can the functions of working memory be incorporated into a non-representational psychology? James Gibson offered some vague metaphorical language which is relevant here—specifically, the suggestion that perceptual systems ‘resonate’ to information [Gibson \(1966\)](#). Presumably resonating is a process that can extend in time. In his later work on visual perception, Gibson also discussed how partially occluded objects could be perceived as whole. The proposal was that perceiving the whole object does not require filling in the occluded part from memory, as had been assumed. This filling in process is not necessary because there is enough information in optic flow for the whole object to be perceived directly: partial occlusion entails that something is being occluded ([Gibson, 1979](#)). A more recent suggestion, from [Pan et al. \(2013\)](#), is that temporally extended perception, or what they call ‘embodied memory’, is possible because events that have occurred and were perceived have calibrated the perceiver’s perceptual system to see the remaining structure in a particular way. The authors give the example of a child peeking while playing hide and seek: the child sees all of the other children disappear behind occluding surfaces, and can now see only those surfaces; but the child’s visual system is now calibrated to see those surfaces as ones that are hiding other children.

Clearly there is much to be worked out here if the aim is to provide an ecological realist account of temporally extended perception. I suspect that the ultimate solution will depend on an assertion that working memory is not ‘memory’—it is not a thing stored on a slate. Instead, we should talk about two processes: ‘remembering’, for the act of reconstructing a past event, and ‘resonating’, for the temporally extended aspect of active perceiving. This second is what is usually referred to as working memory. An account of how to accommodate clarification requests into the ecological realist framework might then go as follows. If S1 produces an utterance, then for as long as S2’s auditory perceptual system is resonating to that utterance, the utterance still exists for S2, and exists as a relation that can itself be acted upon. If S2’s clarification-

seeking action succeeds, it must be the case that S1's perceptual system was also still resonating to the target utterance. This also accounts for why some clarification requests themselves fail (Clark, 1996): if S2 seeks to clarify something which is no longer resonating for S1, then S1 will be confused in turn, and may seek to clarify the clarification request.

5.3 Summary

This chapter has consisted of a further critique of the conduit model, and of an attempt to develop the field model in such a way as to incorporate the main findings that have arisen from the study of referential communication games. It was claimed that the traditional conduit model account of these games—the truncated conduit account—is inadequate because it is based on unrealistic input–output assumptions. A modified, extended conduit account, meanwhile, leads to causal opacity: information flows constantly around an extended system. This raises a methodological problem: it is not clear to what extent the behaviour observed in these studies should be attributed to endogenous properties of subjects' cognitive systems, and to what extent it is ultimately a reflection of structure in the materials placed in the environment by the investigator. It was suggested that the field model does not suffer this problem because the experimental materials are taken at the outset to be part of the phenomenon to be explained.

The field model aims to eradicate language–environment dualism. It aims to provide a basis for a non-inferential account of speaking. The analysis of the structure of referential communication games hinted that there may be a rich enough supply of information in the relational web such that inference is not necessary. This is some work to be done to develop this account into a useful set of practical tools.

Most usefully, the discussion in this chapter has forced us to develop some new ideas about how the field model might be made to work. Affordances can be created through dialogue; this is necessary whenever an object is given a new name. Individuals' relational webs must be calibrated for a given situation. This is achieved through exploration and interaction. Partner-specificity is primary; partner non-specificity is derived. Speaking in monologue is also a derived phenomenon; monologue can be

achieved through the use of special compensatory skills such as practiced, formulaic verbal actions or through the use of special external resources such as written notes. And it appears to be necessary that speaking is phenomenologically extended in time; this has to be so to explain why we can refer to utterances that have already happened.

Chapter 6

The ecosystem of vervet monkey alarm calls

The discussion up to this point has been quite general in nature. It has been preoccupied with the question of what kind of metatheoretical assumptions are necessary if we wish to build a non-representational programme for studying a set of complex social phenomena—collaborating, speaking, learning to act with others. The present chapter will aim to develop this programme in more detail, through the use of a case study of a particular phenomenon. The aim here is to construct an analysis in ecological realist terms of a relatively simple behavioural system: vervet monkey alarm calls. Chapter 7 will attempt to extend this analysis to a human activity. We will look at monkeys first, however. Doing so will help us to avoid invoking propositional content as part of the explanation; we are less inclined to imagine that other species think in terms of sentences and truth conditions.

Vervet monkeys in the wild are observed to emit vocal signals in response to the arrival of a potential predator, such as a leopard or an eagle. These calls have the effect of warning other monkeys of the potential danger. Monkeys that hear the signal will take appropriate evasive action: a leopard alarm call causes monkeys to scatter up trees, an eagle alarm call prompts them to seek shelter. Vervets are one of many species that exhibit alarm calling behaviour; many bird species also make alarm calls, as do ground squirrels, as well as other species of monkey ([Owings and Morton, 1998](#)). There is no particular reason, for present purposes, to focus on vervet monkeys rather than any of

these other species. It will, however, be useful to examine the behaviour of a single species in some detail. Different species' alarm call systems may be very different from one another, in terms of both evolutionary history and developmental origins. This chapter will focus narrowly on vervet behaviour. The aim will be to provide a plausible account of the known facts, using the ecological realist tools introduced in previous chapters. How does this behaviour work? How does it develop? Can it be understood as a system of actions on a relational field?

I will stress at the outset that I make no argument about whether, or to what extent, monkey alarm calls are analogous to human speech. The alarm call system is examined here as an interesting phenomenon in its own right. A reasonable amount is known about alarm calling behaviour in vervets. This is data that we can usefully attempt to fit into the relational field framework. I will comment at the end of the chapter about the extent to which the study of animal signalling is relevant or can inform the study of infant vocalizing and human speaking in general.

6.1 Some facts about vervet alarm calls

Vervet monkey alarm calls have been studied in some detail, notably by Dorothy Cheney and Robert Seyfarth. Their findings from field research conducted over the course of 11 years are presented in their book *How Monkeys See the World* (1990). Their research seeks to address questions about much of what goes on in vervet monkey social life—questions about in-group and out-group behaviour, about differences in behaviour among individuals in a social hierarchy, about deliberate deception, and about the evolution of social behaviour in general. I will focus narrowly on their data on the use of predator alarm calls, and on how young vervets learn to use these calls.

The first fact to consider is that different calls result in different behaviour. Vervets produce a variety of different calls (Struhsaker, 1967). The three main types identified by Cheney and Seyfarth are the leopard call, the eagle call, and the snake call (Seyfarth et al., 1980). These correspond to vervets' three main predators. (Further types of call are made on the sighting of baboons, of humans, and of other mammals that do not typically attack vervets.) Leopards constitute a threat to monkeys that are on the ground in open territory. When the leopard call is made, monkeys that are on the

ground will run up a tree. The tree affords safety from leopards because vervets are small and can easily and quickly move about between the branches. Eagles, however, can attack vervets in trees as well as on the ground. The eagle alarm call prompts vervets to look up, and causes those that are already on the ground to seek shelter in a bush. It also causes some vervets that are up in the trees to scramble down and run for the bushes. Snakes appear to be the least dangerous of the three predators. Vervets do not seek shelter from snakes, but stand bipedally upright and monitor the snake's movement through the grass. A group of vervets will often mob the snake, 'repeatedly giving snake alarm calls from a safe distance' (Cheney and Seyfarth, 1990, p. 103).

Second, there's a developmental trajectory in the way that infant vervet monkeys make calls. Young vervets produce a full range of adult-like sounds from quite early on—from one month of age. But while adult vervets appear to make the three main types of alarm call almost exclusively in response to a sighting of a specific type of predator (the leopard call follows the sighting of a leopard, and so on), younger vervets appear to respond to a high proportion of false positives. For instance, adults make the eagle alarm call in response to only the two types of eagle that prey on vervets (the crowned eagle and the martial eagle), but infant and juvenile vervets are observed making the eagle call in response to non-predator eagles, and also in response to other airborne animals—geese, herons, bustards (Cheney and Seyfarth, 1990, p. 131). Adult-like alarm calling behaviour emerges gradually over the course of the first two years of development.

Third, there appears to be a dissociation between the way that young vervets respond to calls made by adults and the way that they themselves produce calls. This was tested in a playback study in which adult calls associated with the three different predators types were played over a loudspeaker to pairs of mothers and their infants (Seyfarth and Cheney, 1986). Young infants—younger than four months—tended to respond (in more than half of observations) by running to their mother. Slightly older infants were less likely to run to their mother, but also less likely to respond appropriately (e.g. they sometimes responded to a leopard call appropriately by running up a tree, but sometimes responded in a maladaptive way, by running into a bush—a shelter which does not afford safety from leopards). Vervets aged between six and seven months, however, responded in an adult-like fashion in over 80% of observed trials. In

effect, young vervets learn to respond to alarm calls in an adult-like way by the age of six months, although they themselves continue to produce false positives until around age two: ‘In vervet vocal communication, as in [human] language, comprehension precedes production,’ (Cheney and Seyfarth, 1990, p. 137).

Finally, in addition to responding to alarm calls made by conspecifics, vervets also respond to alarm calls made by another species—the superb starling. This starling lives in the same area as the vervets Cheney and Seyfarth observed (Amboseli National Park in Kenya). It produces two types of alarm call of its own: one for aerial predators and one for terrestrial predators. The set of predators for vervets is not the same as the starling’s set of potential predators. Vervets are themselves among the starling’s terrestrial predators, and they elicit the starling’s terrestrial predator call. The starling’s aerial predator call, meanwhile, is made ‘in response to at least eight species of hawks and eagles, only one of which preys on vervets,’ (Cheney and Seyfarth, 1990, p.159). Nevertheless, there is some overlap in predators, which means that the bird’s call is a potentially useful source of information for vervets. And indeed, vervets are observed to respond adaptively to the two types of starling alarm call. In a playback experiment in which starling’s calls were played to groups of vervets on the ground, the starling’s terrestrial predator alarm call caused over half of the vervets to run for trees (but not look up), while the aerial predator call caused over 75% of the vervets to look up (but not run for trees; Cheney and Seyfarth, 1985a).¹

6.2 Alarm calls as actions in the relational field

The four phenomena identified above will serve as useful target material for an attempt to understand vervet alarm calling behaviour using the relational field model. Can these phenomena be explained in terms of actions on the relational field, and if so, how does the relational field explanation compare to Cheney and Seyfarth’s own interpretation of their data?

¹(Cheney and Seyfarth, 1990, p. 160) note that this initial study was carried out on groups containing several juveniles, who are apparently more dramatic in their responses compared to adults. In a subsequent study in which all of the subjects were mature adults the playbacks caused less running for trees; the adult sample were more likely ‘to remain seated and simply scan the area around them’ (Cheney and Seyfarth, 1988).

A major difference to note at the outset is that whereas the relational field model requires that vocalizations be understood universally as actions, [Cheney and Seyfarth](#) are inclined instead to treat alarm calls as labels with distinct referents, analogous to the view of human language as a formal system. Young vervets, they write, ‘behave as if predisposed to divide the events in the world around them into broad categories that require a grunt, a scream, an alarm call, or no vocalization at all. Over time, pronunciation improves and infants sharpen the relation between a call and the objects to which it refers or the context in which it is used’ ([Cheney and Seyfarth, 1990](#), p. 138). Thus, from the assumption that alarm calls are labels, it follows that vervets must have an internal category to represent distinct types of predator. The category is necessary as an explanation of how monkeys are able to respond appropriately to distinct threat types.

Whether or not vervets divide the world into categories is relevant to an explanation of the first phenomenon: how do vervets respond appropriately to different calls? The ecological realist explanation, of course, cannot appeal to an internal category. Instead, the assumption will be that there is in fact no need for an internal category: the business of engaging in alarm calling behaviour is not about forming and using categories; it is about animals perceiving and acting on affordances.

An outline of an explanation of this behaviour in realist terms might go as follows. What a vervet perceives when it perceives a predator is a particular relation—a particular correspondence between itself and a piece of structure in the world (a leopard, say). The presence of the predator is inherently meaningful to a vervet, because predators are things that are disposed to attack monkeys. An adult vervet has learned to attend to the correspondence between itself and the predator. And since this correspondence is always present whenever a predator alarm call has to be made, there is no need to invoke a category: the call is made simply in response to the perceiving of the affordance. When a vervet makes an alarm call, then, it is acting on this very affordance—the immediate threat. How do other vervets respond to this call? Perhaps fellow vervets are themselves *acted upon* quite directly. Hearing the call induces an emotional state in the hearer, which puts the hearer in a certain relation with the environment, even though the hearer has yet to directly perceive the predator. The hearer’s response behaviour is then affected by the way that their relation with the environment has been reconfigured

by the caller's action.

A question that arises is this: when callers call, are they acting on a specific relation? And if so, are they acting on a relation between themselves and the predator (because calling somehow functions to make them less vulnerable to the predator), or are they acting on a relation between other monkeys and the predator—effectively shooing other monkeys to safety? Or perhaps the calls are instead not directed at anything in particular; the calls simply arise as part of a monkey's striving to escape from a particular predator?

Cheney and Seyfarth (1990, p. 145) cite some anecdotal evidence that says that whether or not an individual makes a call is influenced by whether or not other monkeys are present: male vervets apparently do not call when they are alone, out of earshot of the rest of the group, and find themselves ambushed by a predator. This suggests that the calling behaviour is at least sensitive to the present structure of the relational field, although it remains unclear precisely what is being acted on when a call is made.

An interesting case is the snake alarm call. Recall that adult vervets respond to snakes by standing upright and attacking the snake as a group. It looks as though the snake call might be a tool for instigating a type of collaborative activity—collaborative shooing away of the snake. Cheney and Seyfarth speculate about whether this might invalidate their description of calls as straightforward labels: 'as yet, we cannot tell whether a vervet's call should be glossed as a word (simply, snake) or as a proposition (Snake! Let's approach and mob it!)' (1990, p. 174). It should also be emphasized that alarm calls are made not just once, on the first sighting of a predator, say, but are made repeatedly. Vervets continue to call even when all individuals have become aware of the threat. In the case of snake-mobbing behaviour, it is possible that these repeated calls serve as a coordinating device: the calls function to recruit fellows to collaborate in attacking the snake, and they facilitate the act of attacking itself because the repeated calls function to keep track of where the snake is.

However, these repeated calls are made not just for snakes, but also for other predators such as leopards, from which vervets 'simply flee' (Cheney and Seyfarth, 1990, p. 219). Cheney and Seyfarth are perplexed by this, since if calls are simply informative labels, there appears to be no reason to keep repeating them after they have produced

their intended effect. A plausible alternative explanation is provided by Owings (1994), who suggests that vervets' and other species alarm calls are not best understood simply as labels that alert other to the presence of a predator. Rather, the calls also serve a 'tonic' communicative function (Schleidt, 1973), that is, they provoke a temporary change in the emotional state of individual animals, with the effect of making others more alert to their surroundings. The suggestion is that repeated calls function to maintain a heightened state of vigilance in others. The caller benefits from this because any of its fellow vervets who happen to be around to hear the call are in effect recruited into and become part of an 'early warning system' that in turn alerts the caller of further threats (Owings, 1994).

On Owings's view, vervet alarm calling behaviour is understood, in effect, as a kind of functional action carried out by an individual engaged in regulating its own relationship with its environment. This view is quite consistent with the idea that discrete responses to discrete threats can be explained without positing an internal category in the vervet's mind. Instead, the category-like nature of responses emerges as an accidental by-product, one that is observable only by a certain kind of third-party individual, namely a human who has a set of linguistic tools for delineating categories of things. For the vervet, there are simply structures and events in the environment that arise and that it must respond to. It perceives how to respond by discriminating the properties of these events.

Clearly, to be able to make and react to alarm calls appropriately, individual vervets must either learn to fear things that are likely to cause them harm, or else they must be predisposed to fear such things. To explain how adult vervets come to be discriminating perceivers of discrete threats, it will be necessary to examine the second phenomenon—the developmental trajectory of alarm calling behaviour. The ecological realist explanation for this developmental trajectory must be consistent with the doctrine of the new ball. A young vervet must be *changed*, over the course of its development, such that it responds more selectively to actual threats. The working assumption must be that there is sufficient information available to the young vervet as a result of its own exploratory activity that it is able to accurately learn what the real threats are. There must be structure that a young vervet can attend to that allows it to learn to discriminate the threats from the other, incidental events going on in its

environment.

Why might infant vervets call for things that are not real predators? [Cheney and Seyfarth](#) assume that this behaviour is simply a ‘mistake’ (usually with the scare quotes; [1990](#), p.129): it occurs because the infant is in possession of a set of not-yet-fully-refined predator categories. But if calls are actions, and not labels for categories of predator, then this explanation does not make sense. According to the pragmatist philosophical tradition, perceptions and actions cannot be true or false—they cannot be mistaken—because in order to be so classified they would have to exist in the form of propositions, which they do not ([Michaels and Carello, 1981](#), p. 108). A ‘mistake’, like a ‘category’, is something that can only be identified from the outside, and with the use of a linguistic description scheme.

[Cheney and Seyfarth](#)’s assumption is that young vervets are simply badly-performing or less-than-fully-competent adults. [Owings \(1994\)](#) again draws this into question. [Owings](#) suggests that it may be a mistake to try to understand infant behaviour in terms of an adult target. Perhaps the young vervets are not striving to behave like adults, but are behaving in a way that is appropriate for their specific developmental stage. It is possible that an infant’s alarm calls, or at least its ‘mistakes’, may be best understood not as attempts to alert others to a danger, but as exploratory actions carried out by the infant as it attempts to assess what the environment affords for itself. [Cheney and Seyfarth \(1990, p. 133\)](#) themselves provide some reason to think that this is the case. They note that the way that adults respond to infants’ calls is a potentially important source of information for the infant. If the infant calls in response to a goose, then there is no real threat; adults will disregard the infant call and go back to what they were doing. However, if the threat is real—the infant call was caused by a martial eagle, say—then adults are likely to produce their own call in turn, echoing the infant and confirming that the call is appropriate. Thus, infants’ calls function in part to elicit adults responses, which are a potentially useful source of learning. As [Owings \(1994\)](#) points out, this infant-initiated learning strategy makes sense, because adult vervets do not appear inclined to deliberately teach their offspring.

If this exploratory learning hypothesis is correct, it provides a mechanism through which young vervets can come to discriminate structure in energy arrays that specifies the predators in their environment. Importantly, it constitutes a kind of exploration

that is not likely to endanger them personally. When a young vervet perceives a flying object which (to it) is ambiguous, its alarm calling behaviour is a means of generating further information about what that object affords. The infant's behaviour is analogous to the partner's behaviour in the tangram game, discussed in the previous chapter: non-speaking overhearers are disadvantaged relative to speaking partners because the partner can generate useful information which an overhearer cannot (Schober and Clark, 1989). A prediction that follows is that young vervets that are somehow prevented from making their own calls would be impaired in their ability to discriminate threats in an adult-like way. Perhaps mute vervets are slower to arrive at adult-like behaviour, or perhaps they never arrive there, but instead continue to be agitated by non-predator eagles into adulthood.

This explanation may appear to leave unexplained the third phenomenon: why is there a dissociation between the calls young vervets make (adult-like at 2 years) and the calls they respond to (adult-like at 6 months)? As mentioned in chapter 3, a similar phenomenon occurs in human infancy: infants respond appropriately to a greater number of words than they themselves produce (Benedict, 1979; Goldin-Meadow et al., 1976). It was suggested that this is not a problem for the relational field model: this discrepancy becomes a puzzle only if it is assumed that 'language' is a single internal system—a system arranged around an internal lexicon, an internal set of grammatical rules, and so forth. From the infant's point of view, however, the verbal actions of others are simply a set of structured events, and the things an infant tries to achieve are a different set of events for which appropriate verbal structures have to be learned at first hand. The dissociation in vervets is no more a problem than it is in humans. If infant verbal behaviour can be explained without positing an internal system of language, then vervet alarm calling behaviour can be explained without invoking an internal system of signals. The calls of others are events in the world. Learning the meaning of these events is not the same as learning to cause similar events.

The dissociation is even less puzzling if we consider the richness of information available for the two tasks: responding to calls versus emitting calls. Hearing an alarm call is not an event that occurs in some separate decontextualized realm. The infant is not typically isolated from the rest of the group; its task does not consist of interpreting the meaning of a call from an impoverished set of cues. Alarm calls are heard in an

environment populated by other members of the group. These other group members themselves respond actively by fleeing and further monitoring the environment. The infant's responding to the call is therefore also an act of responding to the behaviour of others. A very young infant, in fact, does not necessarily need to interpret the meaning of the call itself—the infant can respond to any call by fleeing to its mother. It will learn to respond appropriately simply by following the mother and observing what she does. Slightly older vervets, furthermore, can assess how to react to a given call by first looking at how an adult is responding. [Cheney and Seyfarth \(1990, p. 136\)](#) present data suggesting that this is a strategy that young vervets use: those that look to an adult before acting nearly always respond appropriately to a given call; those that do not look to an adult are more likely to respond in an inappropriate way—running to hide in a bush in response to a leopard call, say. In short, there is rich information available to an infant about how to respond to a call. Contrast this with the information that is available about an ambiguous flying object that an infant has seen but that no other vervet has yet called for. Information about the flying object is not simply available, but has to be self-generated through the infant's exploratory calling action.

An even clearer demonstration of the discrepancy between responding to and producing alarm calls is provided by the fourth phenomenon: vervets' use of the predatory alarm calls of starlings. Here, vervets make use of (they 'comprehend') a piece of structure that they themselves will never produce. Vervets never become competent producers of starling vocal signals, and nor do they strive to be. Responding to and producing alarm calls are two distinct activities; in [Owings and Morton's \(1998\)](#) terms, the first is assessing the environment or monitoring it for threats and opportunities, the second is managing the environment—attempting to change the structure of what the environment affords in the caller's self-interest.

For [Cheney and Seyfarth](#), vervets' use of starling alarm calls requires a substantial degree of inference-making on the vervet's part: 'For a monkey to learn to respond appropriately to the starling's alarm calls, she must first learn which of the starling's vocalizations function as alarm signals. She must then learn to distinguish between the two types of alarm calls and recognize which predator species evoke the calls.' ([Cheney and Seyfarth, 1990, p. 159](#)). On the relational field model, this is simply not necessary. From the perspective a given vervet, the starling's predator call is simply

a part of an extended event. The starling's call sometimes precedes or accompanies the appearance of one of the vervet's predators (but not always: the starling may be calling in response to an animal that is only a threat to starlings). The occurrence of the starling's call is perhaps similar to a change in the background music on a film: film directors often include portentous music as a way of foreshadowing a dramatic event, or of suggesting the possibility of such an event. It is plausible that the starling's vocalizations might have a similar effect on vervets. Unlike someone watching a film, a vervet can respond to this change in background music by engaging in exploratory activity: assessing the sky or the ground for potential threats. Young vervets may be particularly stimulated by this change in music because they have learned that this change sometimes accompanies danger, but they have not yet learned to respond to it in a discriminating way.

6.3 Monkey alarm calls and human speaking

The examination of monkey alarm calls undertaken in this chapter has made no commitment on the nature of the relationship between this behaviour in monkeys and the evolutionary origins of speaking in humans. I do not intend to speculate about any such relationship here. It will, however, be useful to make some comments about the parallels between the two in terms of their underlying machinery, as a means of judging the extent to which it is appropriate to generalize conclusions drawn from the study of one system to an understanding of the other.

This comparative project is, indeed, a major motivation for [Cheney and Seyfarth's](#) work. They open their book with the suggestion that studying 'almost minds'—things like monkeys, computer programmes, and human infants (p. 3)—is a useful means of studying adult human cognition. In their concluding comments, they discuss what it is that makes monkeys and humans different. The main difference, they conclude, is that humans have a capacity to attribute mental states to others—a capacity that monkeys lack (p. 312; emphasis in original):

The inability to examine one's own mental states or to attribute mentality to others severely constrains the ability of monkeys to transmit information, deceive, or feel empathy with one another. It also limits the extent

to which monkey vocalizations can be called semantic. True, calls function to denote objects and events in the environment and, like words, are caused by the mental states of those who use them. Unlike our language, however, the vocalizations of monkeys are not given with an intent to modify the mental states of others. Though monkeys are skilled observers of each other's *behavior*, they seem to be far less astute observers of each others' *minds*, and they seldom seem to proceed beyond other animal's actions to analyze the motives underlying their behavior. We attribute motives, plans, and strategies to the animals, but they, for the most part, do not.

Evidently, this passage makes an important assumption that is rejected by the relational field model: that human speaking is primarily about altering the mental states of others. On the field model, speaking is instead understood as acting on relations; it is about regulating or managing the behaviour of others relative to the rest of the environment. But the concept of acting on relations serves the same function as [Cheney and Seyfarth](#)'s concept of manipulating others' minds: both are attempts to explain how a speaker directs another individual's attention to a specific piece of external structure.

We are assuming that human speaking is acting on a relation. The question that follows is one that was raised above: is vervet alarm calling behaviour also acting on a relation, or is it something else entirely? Is it a deliberate attempt by an individual to reduce the threat that a predator poses to its fellows, say? In this case, a call would be an action on the relation between other vervets and the predator. Or is the alarm call more like human infant crying, which, it has been suggested, may be better understood as a means by which an infant regulates its own needs ([Thompson et al., 1996](#); [Owings and Zeifman, 2004](#))? In this second case, the call is not an action targeted at a specific relation between another individual and some structure in the environment, but is part of an attempt to deal with a specific, personal threat. In other words, it is an action on a relation between the *caller* and the threat.

There is reason to believe that this second view is correct. For instance, vervets are apparently most likely to call to predators from which they themselves are most at risk ([Cheney and Seyfarth, 1990](#), p. 219). If calling was simply about alerting others, we might expect that calls would be equally likely whenever a predator is present, or would be most likely when others are most at risk and therefore most in need of being alerted. Actually, it's possible that there's a functional, group-level reason why

the monkey calling most loudly is the one closest to the threat. This might serve as a means of localizing the threat for others: it relieves others of the burden of having to look around and scan their entire environment in response to the call. Instead, they can simply look in the direction where the noise is coming from, and the predator will likely be close to the source of the noise. But there are other reasons to think that alarm calling is self-regulation. Infant vervets are vulnerable to attack by baboons, but adult vervets do not appear to increase their production of baboon alarm calls when there are infants present, even though this would presumably benefit their offspring (p. 219). Adults also make no special attempt to correct infants' 'mistakes'; they do not reward appropriate calls, and are no more likely to echo an appropriate call by an infant than they are to echo a call by another adult, even though this echoing may be especially informative to the infant (p. 225).

Perhaps the crucial test case is how adults respond to snakes: how do vervets coordinate an act of snake-mobbing? This, more than any of the other vervet behaviours that have been reported, looks, on its face, like a clear instance of a collaborative phenomenon. It looks like an adult vervet's emitting of a snake call is partly a means of initiating a collaborative mobbing episode. More data is required to assess whether this is truly the case. What's needed is data about how adult vervets respond to the appearance of a snake in different conditions: when there are no other individuals around; when there are only infants around (presumably infants are not yet capable of joining in in mobbing the snake); when only one other adult is present; and when multiple adults are present. If the snake call is really an action on a relation between the snake and another individual, then one would expect that adults should only attempt to initiate a mobbing episode when an appropriate team can be assembled. Granted, engaging others in mobbing the snake is also a type of self-regulation: it is a way of dealing with a personal threat. But it is also an activity that requires the monkey to attend to the kind of triadic structure that—it was suggested in chapter 1—is characteristic of the things we think of as collaborative activities.

In the end, it remains unclear whether vervet alarm calling is essentially a means of self-regulation, or whether it is ever a case of acting on others' relational webs, and thus analogous to human speaking. That is, we can conclude that it is plausible to describe alarm calling behaviour in terms of actions on the relational *field*, but it remains

an open question whether the calls are directed at first-order relations (at regulating the affordances between the calling individual and environmental threats), or at second-order relations (at relations *between* entities in the environment). The extent to which the study of alarm calling is directly relevant to an understanding of the development of speaking in human infancy depends on the validity of the analogy here. In terms of methodology, however, we can take the discussion in this chapter as a proof of concept that it is possible to study in non-mentalistic terms behaviour that looks ‘communicative’, and that is engaged in by groups of primates who in some ways appear to be like ourselves.

6.4 Summary

The analysis of vervet monkey alarm calling behaviour developed in this chapter contrasts clearly in some respects from the analysis developed by [Cheney and Seyfarth \(1990\)](#). An important conclusion is that the ecological realist approach can deal with an activity that apparently involves categorization, and can do so without positing an internal category. Affordances are correspondences between animals and structure in the world; perceiving these correspondences means perceiving meaning directly, without having to consult any kind of internal store.

Next, there is further reason to think that animal communication cannot be understood simply in terms of information exchange. Dealing with problems is a temporally extended biological activity. Escaping predators requires not that animals simply respond once to a given stimulus source; instead, the animals must actively regulate themselves and others in a fashion that allows them to continue monitoring the environment until the danger has passed. [Owings \(1994\)](#) suggests that repeated alarm calls should be understood as tonic communication which regulates monkeys’ emotional states with the function of maintaining vigilance; vocal repetition is not simply redundant echoing of information.

Young vervets’ production of alarm calls in contexts that do not elicit such calls from adults was taken as further evidence that learning is an exploratory process ([Gibson, 1988](#)). The dissociation between how animals respond to calls and how they themselves produce calls is explained by the fact that learning to act on the relational

field is different from learning to respond to changes in the structure of the field. Finally, while vervet alarm calling can be understood as acting on relations, it is not clear whether alarm calling is ever acting on others' relational webs as such, or if it is self-regulation. Nevertheless, the relational field model provides a useful guide for constructing relevant research questions.

Chapter 7

Action perturbation in a collaborative assembly game

We have just looked at a behavioural system observed in monkeys, and have attempted to analyse this behaviour in terms of actions carried out on the relational field. It might be argued, in response, that such an analysis does not provide a very strong reason to take the relational field model seriously. ‘It’s all well and good to describe monkey behaviour without invoking mental content,’ a hostile observer might say, ‘but what about actual speaking behaviour carried out by cognitively sophisticated humans? Can the relational field model provide a novel understanding of some particular collaborative activity carried out by people?’

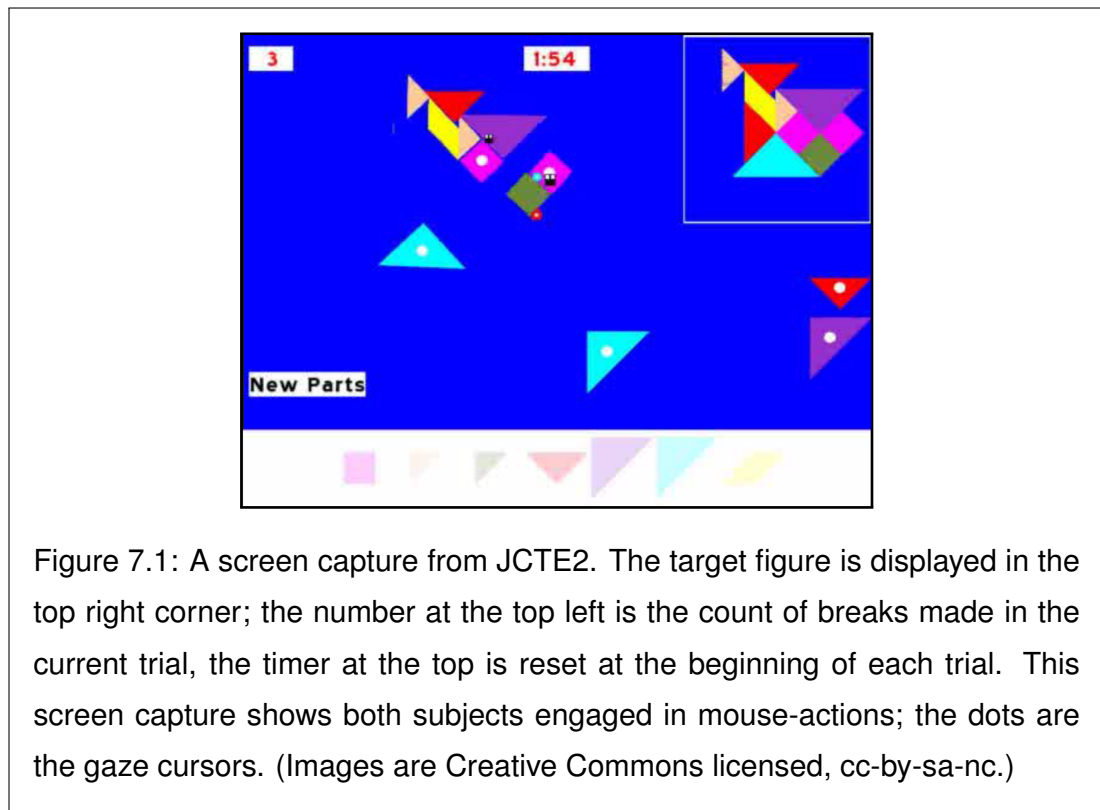
This chapter is an initial attempt to address this challenge. It will examine collaborative behaviour in a referential communication game: the joint construction task (JCT; [Carletta et al., 2010](#); [Bard et al., 2014](#)). The data discussed here are drawn from a pre-existing corpus. This task differs from the standard referential communication game format discussed in chapter 5. In the JCT, there is no barrier; that is, the two subjects are not prevented from seeing one another’s materials. Instead, both subjects see exactly the same set of materials throughout the session. The subjects are each sat in front of their own computer monitor, but the same objects are displayed on both monitors simultaneously, and both subjects can freely rearrange these objects in real time. Experimental manipulation in the JCT consists of alterations that are made to the set of tools the subjects have at their disposal for interacting with the objects dis-

played on their screen. Different experimental conditions manipulate whether or not the subjects can speak to one another, whether or not they can see one another's mouse cursor, and whether or not they can see their partner's gaze cursor, which is captured via a head-mounted eye-tracker.

It was argued in chapter 5 that the standard referential communication game format, complete with the barrier, is not a suitable tool for an empirical programme that conceives of speaking as acting on a relational field. The typical referential communication game is constructed on the presumption that speaking is transmitting or sharing information through a conduit. The JCT is built on a similar presumption, but it is nevertheless a more interesting activity for present purposes. Because there is no barrier, it is not the case here that an artificial conduit is created by the format of the game itself. However, the game is presented to subjects deliberately as a novel activity. Because of this, it will be necessary to exercise some caution about what the task in fact is, from the first-person perspective. This will be discussed in more detail below.

A game trial in the JCT is set up as follows. A pre-assembled target model is displayed in the top-right corner of the screen (see Fig. 7.1). Underneath, a set of pieces are laid out. Subjects are instructed to use these to reconstruct the target figure. The game software is programmed in such a way that the figure can only be built if the subjects collaborate in certain ways:

- Both subjects can freely move pieces around the workspace, by clicking and dragging them with the mouse. Each subject controls a single mouse cursor.
- In order to join two pieces together, each subject must click and hold one of the pieces as the two are brought into contact. The resulting larger piece can then only be moved around as a single unit.
- If both subjects click on the same piece at the same time, the piece breaks (it disintegrates on the screen and immediately appears again in regenerated form in a special area at the bottom of the screen).
- A piece will also break if a subject brings it into contact with the edge of another shape that is not currently being held by their partner.



Notice that the way the game is set up means that some of the actions that subjects can carry out are both directly visible to their partner and inherently meaningful in that the action carried out immediately alters what the game environment affords to the partner. Most obviously, being able to see the other person's mouse cursor is potentially a useful source of information about what actions are currently afforded in the game. When S1 clicks on a piece, that piece now affords breaking by S2. It also affords joining another piece on. A piece that has not been clicked on by the partner, meanwhile, affords being clicked on and dragged around the screen.

There is thus reason to expect that whether or not the partner's mouse cursor is visible will impact on how well dyads are able to perform. It is less clear what we should expect regarding the other factors that are manipulated. Consider the partner's gaze cursor. This is not inherently meaningful in the same way as the mouse cursor. Suppose that S1 can see that S2 is looking at a particular piece. This does not in itself specify to S1 anything about what actions that piece currently affords. Pieces break when they are clicked on by each subject, but not when they are merely looked

at. In addition, the gaze cursor jumps around the screen in step with the partner's eye-movement saccades. This is an alien kind of movement that we do not encounter except when viewing videos produced using eye-tracking equipment; it is a kind of movement that initially looks very odd. So while the gaze cursor is certainly a potential source of information for a subject in the JCT, it is one that they will likely find is difficult to make use of.

The third factor that is manipulated is whether or not the dyad can speak. Speaking is not an essential part of being able to assemble the figures, as it is in the card-sorting games where the partner's cards are hidden behind a barrier. It is not, therefore, immediately clear what role speech might play here, or whether being able to speak to each other will necessarily lead to improved performance. In the experiments in the JCT corpus, each of these sources of potential information is manipulated systematically in various ways. These manipulations are summarized in Table 7.1.

As can be seen in the table, the experimental manipulations of interest to us here were all presented within- rather than between-dyads. This means that each dyad had to complete a series of 16 trials under a set of conditions that varied as the session progressed. And each set of conditions was presented in a block, of four trials in length (in JCTEs 1 and 3), or of two trials (in JCTE2). The JCT was set up in part, like many of the studies discussed in chapter 5, in order to investigate how people make use of referring expressions in dialogue: how subjects form 'conceptual pacts' with their partner, whether they take into account their partner's perspective, and so on. The trial-order was counterbalanced between subjects in order to mitigate possible confounding effects of order. As will be seen, however, this shuffling of the conditions within each experimental session created some interesting patterns in the data. A number of different measures that we can look at in the data show evidence that dyads are learning as the session progresses. And when we look specifically at the points in the session at which a new set of conditions is introduced, we can clearly see that this disrupts the general learning trend (e.g. the first trial under a new set of conditions exhibits a slower completion time relative to the previous trial); that is, changing the environment *perturbs* the dyad's action pattern. In the present chapter, I will examine the JCT corpus data as if the game had been set up to investigate these perturbation effects in learning. It will be useful, first, to elaborate slightly on the concept of perturbation.

JCTE1	JCTE2	JCTE3
4 conditions; each presented in a block of 4 trials	8 conditions, presented in 2-trial blocks	4 conditions, in 4-trial blocks
Target figures composed of 7 pieces all of different colours	Target figures made of 11 pieces, some of which are duplicated	7-piece target figures; 5 pieces are each a different shade of green
Pieces provided: the same 7 pieces as in the target figure	13 pieces provided, two of which are not needed for the target figure; there are 6 pairs of duplicates and 1 unique piece	Pieces provided as in the target figure
Speech manipulated: dyads can speak in only trials 1–8 or in only trials 9–16	Speech manipulated as in JCTE1	Subjects can speak throughout the session
Gaze manipulated: dyads can see each other's gaze cursor in 8 of 16 trials	Gaze manipulated as in JCTE1	Gaze not manipulated; dyads do not see each other's gaze cursor
Mouse cursor not manipulated; dyads can see each other's mouse cursor throughout	Mouse manipulated: dyads can see each other's mouse in 8 of 16 trials	Mouse manipulated: subjects can see each other's mouse in 4 trials, cannot see each other's mouse in another 4 trials; and the remaining 8 trials are asymmetric—one subject can see their partner's cursor but the partner cannot see theirs

Table 7.1: Differences between JCT experiment corpora 1, 2, and 3

7.1 Learning and perturbation

By now perhaps the standard example of perturbation in learning is provided by Esther Thelen and Linda Smith's work on the so-called A-not-B error in human infants (Thelen et al., 2001; Thelen and Smith, 1994; Smith et al., 1999). This is a phenomenon originally identified by Piaget which is observable in young children, around 10–12 months of age. The basic apparatus for this experiment consists of a toy, and a board on which are placed two containers in which the toy can be hidden. The toy is first visually presented to the child. The experimenter then hides the toy in container A, and pushes the board towards the child, whereupon the child is allowed to retrieve the toy and play with it. This procedure is repeated a few times, and then there is a crucial trial in which the hiding place for the toy is switched, and the experimenter now places the toy in container B. This hiding event is followed (crucially) by a delay of a few seconds. Then the board is again pushed within reach of the child. The 'error' is that the child will now reach 'perseveratively' for container A once again, even though the child has observed the toy being hidden only seconds before at the new location, B.

Piaget took this error as evidence that children lack a fully-developed concept of objects as things that, once hidden, remain in the same place. Subsequent studies, however, revealed that the likelihood that a child will commit the error is sensitive to a great range of contextual factors, such as how many containers there are, whether these are transparent or opaque, how long the delay is in the test trial, whether the target object is a toy or a piece of food, whether the test is conducted at the child's home or in a laboratory, and on and on (Thelen et al., 2001). Clearly, Piaget's explanation is inadequate to account for all of these effects. Subsequent attempts to explain the phenomenon have appealed to such things as processing demands, or to a conflict between egocentric and allocentric representations, or to infants' underdeveloped brains as possessing a reduced ability to inhibit the motor response causing them to reach to A. Each of these failed in some way to account for the full range of observed effects.

Thelen and Smith's solution conceives of the child as a dynamic system, whose behaviour is shaped by its past experience. They reject all talk of an object concept (which is a kind of internal content; it is symbols on a slate) in favour of a view of the child as a system that changes over time (i.e., learning-as-change, or the doctrine of the

new ball). In the jargon of dynamic systems theory, repeated behaviour by an organism is said to lead to the formation of an attractor, or basin of attraction. This rests on a metaphor that pictures the set of possible behaviours available to the child as forming a landscape with various peaks and troughs (or basins). The child's actual behaviour can then be represented as a ball which rolls around this landscape. When the child is put into a familiar-looking situation, the child's behaviour naturally 'falls into', or is 'channelled' by the existing attractor basins. If the child's previous reaching experience has formed a strong basin of attraction for reaching-to-the-left-in-response-to-this-specific-apparatus, then the A-not-B 'error' will naturally arise whenever conditions are such that no other, stronger, attractor interferes with this response.

The shift of the hiding location to B is a perturbation of the infant-environment system. A specific response has been established within this system through the practice trials: reaching to A. The perturbation means that this already-established response no longer corresponds to the facts about where the object is hidden. In order for the child to overcome this switch, the system must be re-organized in some way. The point, for present purposes, is that these perturbation effects are informative about the way the system has come to be arranged. If we can interfere with the environment in some way, and this has an effect on behaviour, this constitutes evidence that the thing interfered with was instrumental in how the task was being carried out. If, by contrast, we interfere with something and the behaviour carries on as before, this suggests that the thing interfered with was incidental to the task.

To put this in terms of the relational field: learning consists of attending to some particular part of the environment. Perturbation occurs when that affordance relation is made inaccessible, through, in the case just described, an alteration that happens to the environment (the toy moving to B). Note, also, that a perturbation may equally occur through a change to the animal which affects its ability to act relative to the environment, such as through brain injury, or through a temporary change in the animal's capabilities (e.g. when a person is carrying a large box), or simply through learning (recall the discussion of the transition from crawling to walking in chapter 4).

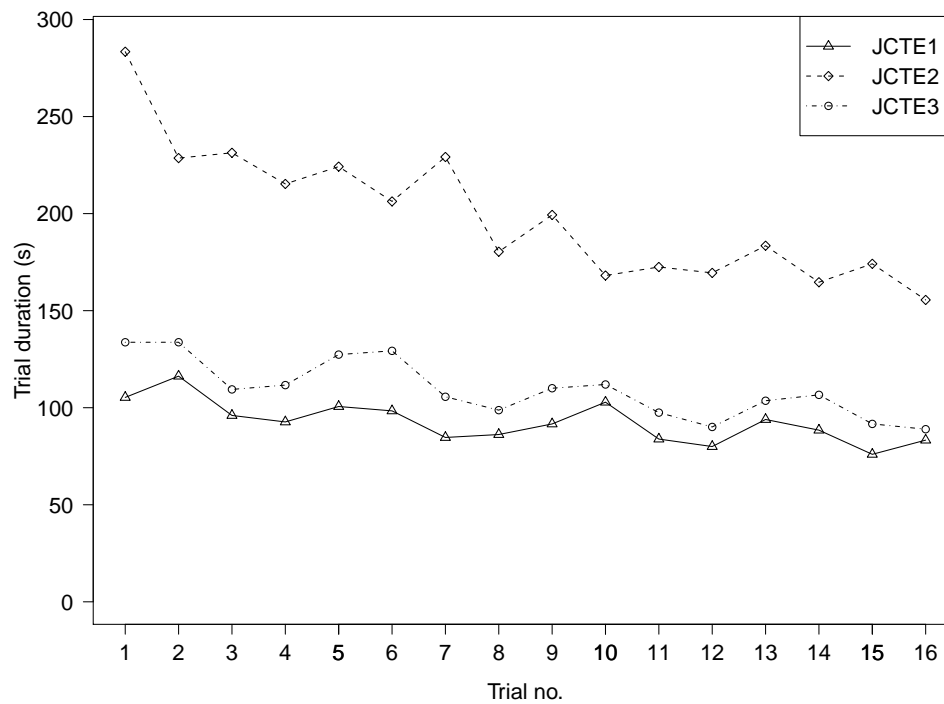


Figure 7.2: Mean trial duration, plotted over the course of the session for each of the experiment corpora

7.2 Perturbation in the JCT

Perturbation effects, so defined, are readily observable in the JCT corpus data. We can break the data down by trial number and look at what happens to various performance measures as the session progresses. See the graphs showing a sample of these measures: the overall length of the trial in seconds (Fig. 7.2), the total number of words spoken by the dyad during a given trial (including only trials where the dyad were allowed to speak; Fig. 7.3), and the number of mouse-actions carried out in assembling the target figure (Fig. 7.4). There are two things to remark about these: 1) performance steadily improves as the session progresses, although 2) at each point in the session where the game conditions are shuffled, performance tends to worsen slightly relative to the immediately preceding trial (in JCTE1 and JCTE2, this happens after every four trials, while in JCTE2 it happens after every two trials; this is because there

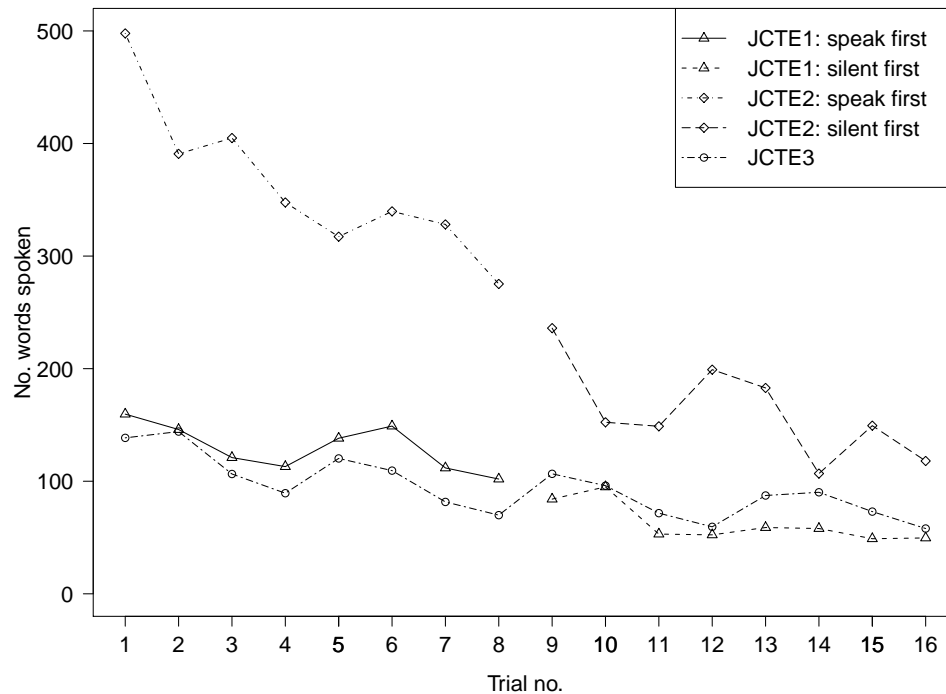


Figure 7.3: Mean number of words spoken per trial. In JCTE1 and JCTE2, pairs could only speak in 8 out of the 16 trials, either trials 1–8 ('speak first') or trials 9–16 ('silent first'); in JCTE3, all pairs were allowed to speak throughout the session

are twice as many sets of conditions in JCTE2, as detailed in Table 7.1).

The claim here is that these results can be explained in terms of perturbation of the perception-action relations between the subjects and the materials. In the first trial in a new block, the tools that are available to the subjects (mouse, gaze, speech) have been shuffled relative to previous trials. This means that the pair must adjust to these new conditions. By contrast, in the second trial in that block (and also in the third and fourth trials in a block, in JCTEs 1 and 3), the conditions are held constant with the immediately preceding trial. This means that in the later trials in a block, the pairs have had the opportunity to settle into or establish an appropriate pattern of action. We can clearly see this perturbation effect in Fig. 7.5, which shows how the learning effect that is observed over the course of the experimental session (in JCTE2) is distributed in an uneven but systematic way. The 'baseline' figure is the mean increase in speed over the

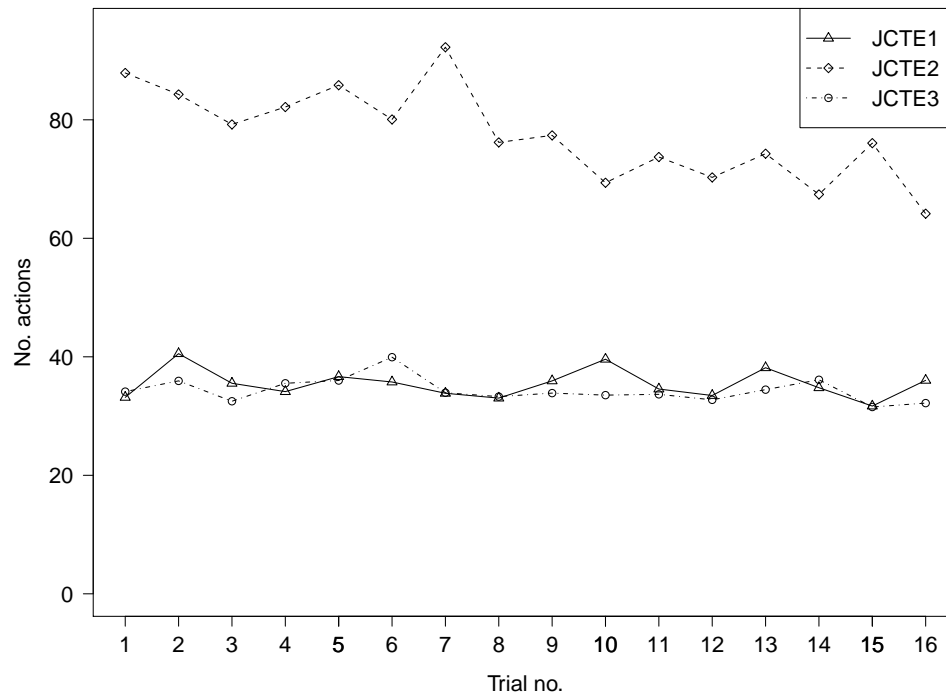


Figure 7.4: Mean number of mouse actions performed per trial

course of the entire session, calculated as the difference between the durations of trial 1 and trial 16, divided by 15 (for the 15 steps between trials). Trial-to-trial learning effects are then calculated for trials 2 through 16, as the increase in speed relative to the immediately preceding trial. Fig. 7.5 compares this learning effect in the first trial in a new block relative to the second trial in a block. Clearly, pairs are slowed down considerably in the first trial in a new block, and show a considerable increment in speed in the second trial.

However, Fig. 7.5 does not show the reason for this pattern. What is it about the first trial in a block that slows the dyad down? Are they slowed down by the mere fact that the conditions have changed, or is there some particular manipulation that is driving this pattern? Here, the existing data set for the JCT is not able to provide clear answers. One way to assess the relative effects of the different manipulations would be to look specifically at the very first trials where a given tool (mouse cursor, gaze cursor, or speech) is either introduced or removed, for a given dyad. However,

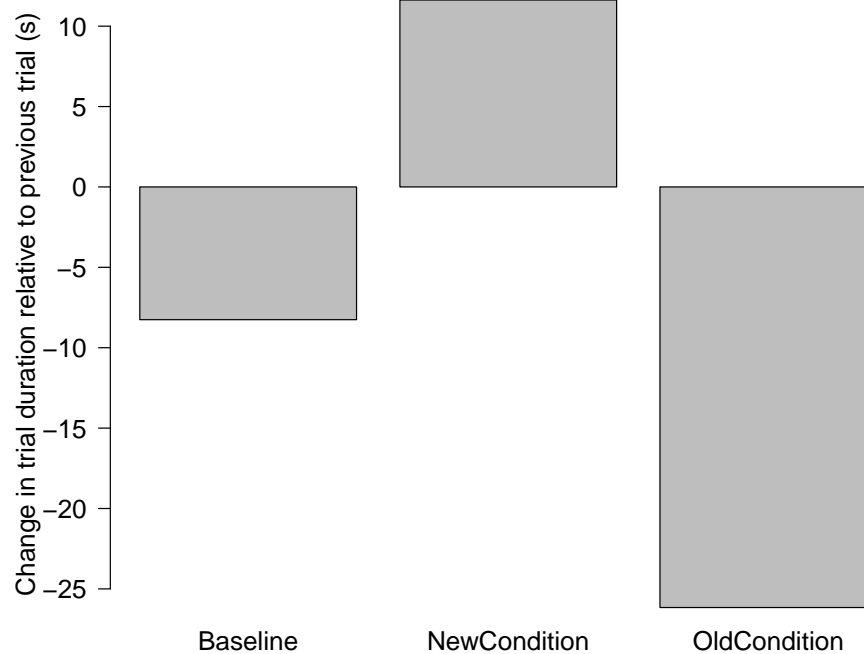


Figure 7.5: Trial-to-trial change in completion time in JCTE2; baseline = mean change across the whole session; ‘new condition’ trials are the odd numbered trials (excluding trial 1), ‘old condition’ are the even numbered

because of the way the experiment was set up, these manipulations are all confounded with one another. And in order to compare individual manipulations within a given corpus (e.g., to compare trials in which the gaze-cursor is added for the first time with trials where it is removed for the first time), it would be necessary to divide the corpus into two groups of 16. But there are substantial individual differences between dyads which means that with an N of 16 the noise-to-signal ratio is too high to allow us to be confident about drawing any reliable conclusions from the data set.

Nevertheless, it is clear from the figures, and especially Fig. 7.2, that the perturbation effect is real, and that it reliably occurs, across all three corpora, whenever dyads transition from an old block to a new block with a new set of conditions.

7.3 What is the task in the JCT?

I have argued elsewhere that in order to study a task empirically, it is necessary to know what the activity looks like from the first-person perspective of the person carrying it out (chapter 4; Baggs, 2013). While we can be fairly confident about what the individual is trying to do in certain well-practiced tasks such as ball-catching, things are less clear for novel tasks such as in the JCT, where subjects are deliberately asked to carry out an activity with a set of materials that they have not previously encountered. It will be useful to ask the following questions about the JCT. What is the task, from the point of view of the subjects? And of the measures: are they measuring what we think they are measuring?

We will consider first the measures already mentioned above. Fig. 7.2 displays mean trial duration across the session. Are dyads really trying to minimize the time it takes to assemble a figure? There is evidence in the corpora which suggests that, for at least some of the subjects, minimizing the completion time is in fact not something they are particularly concerned about. For instance, one pair (dyad 31 in JCTE2), in a trial quite late on in the session (trial 14), spend over four minutes building a figure which they then decide is not accurate enough. One subject is heard saying: ‘Oh that’s awful... you know something, I don’t like it’; he then scraps the assembled construction and the pair start building it again from scratch. Clearly, for this pair, minimizing completion time is at best a secondary concern; they are more interested in maximizing their accuracy score. And in fact, this kind of behaviour is explicitly encouraged in the instruction sheet that was given to the subjects at the beginning of the session: ‘Remember that the success of each joint construction project will be judged on overall efficiency. Efficient trials are fast and, more importantly, accurate; they are low on mistakes that cost breakages, replacement parts and extra building time.’ So we might worry that, if the subjects themselves are not *trying* to reduce their completion time, then it is not clear what the trial duration measure is measuring. In response to this, I will simply appeal to Fig. 7.2 again. Clearly, the completion times consistently decrease as the session progresses (although of course there are substantial differences between dyads for any given trial), and this itself suggests that subjects are indeed, in general, trying to complete the figures quickly.

Fig. 7.3 displays a count of the words uttered during a given trial. Subjects are probably not deliberately attempting to minimize this number. They were not instructed to pay attention to the number of words they were using. In general, the number of words spoken appears to correlate quite closely with the length of the trial, as one might expect (compare the lines for JCTE3 in Figs. 7.2 and 7.3). An interesting feature of the data here is in the comparison between the two sub-groups in JCTE1 and JCTE2: those that were allowed to speak only in the first eight trials and those that could speak only in trials 9–16. The number of words spoken by the latter group in a given trial is always lower than the number of words spoken by the former group even in their least talkative trial. A naïve prediction might have expected the latter group's word count in trial 9 to be similar to the former's in trial 1, because the first trial with speech is the one in which dyads first try to establish a set of referring expressions, and so on. That this is not the case is further evidence that the language used is inseparable from the actions being performed. That is, it is another reason to reject language-environment dualism.

Fig. 7.4 displays a count of the mouse actions performed in a given trial. Again, this is probably not something that the subjects were deliberately trying to minimize. And in fact, efficiency in terms of completion time may not equate to a minimal number of mouse actions. For instance, it may be quicker to move all the pieces into roughly the right position first before joining them all together, even though this involves a greater number of mouse-clicks relative to simply dragging the pieces directly out of the starting area and immediately joining them to other pieces. For this reason, I will not make any strong claim about the meaning of Fig. 7.4. Qualitatively, the figure appears to show that there is a perturbation effect on the number of actions performed in JCTE2 (the line tends to go up to the odd trials and down to the even trials), but no such effect is apparent for the other two corpora. Perhaps in JCTEs 1 and 3 the target figures are sufficiently simple that no learning is necessary, mouse-action-wise, or perhaps there is a learning effect but it is being masked by something else.

What about the task? What is it that subjects are in fact doing as they work through a session of this game? The instruction sheet includes the following: 'Your job is to replicate the model as accurately as possible whilst simultaneously minimising waste both of time and of component parts.' On the face of it, the project of assembling

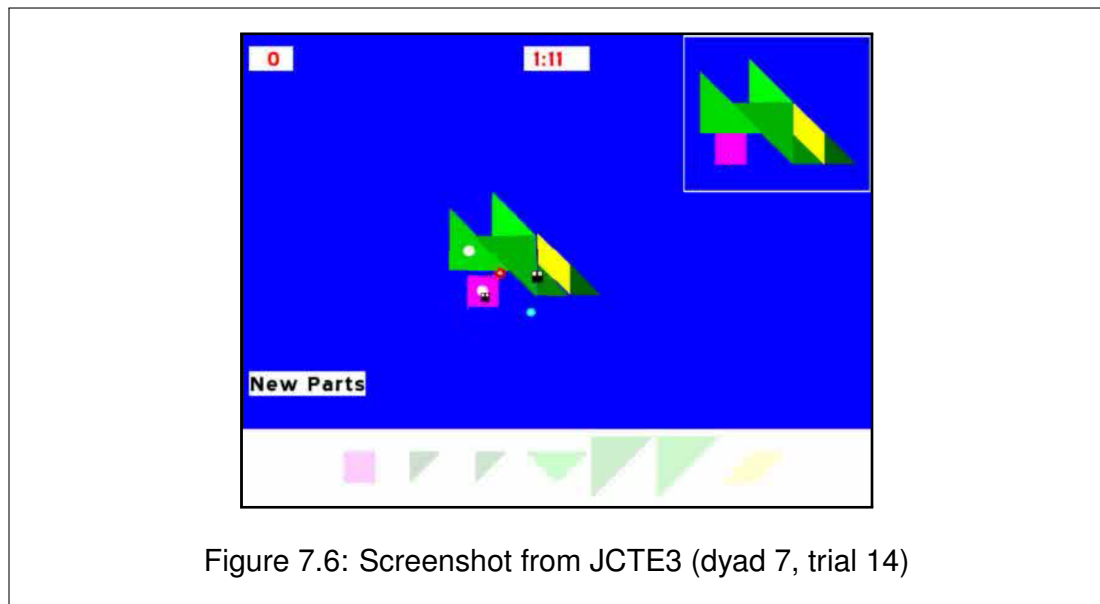


Figure 7.6: Screenshot from JCTE3 (dyad 7, trial 14)

these figures seems entirely meaningless. Yet the data suggest that subjects remained motivated and engaged throughout. At the end of each trial, the software computed an accuracy score for the assembled figure, based on the number of pixels that would overlap if the assembled figure were to be superimposed on the target figure. This score was also displayed to the subjects. I have not included a graph of these scores here because the graph would not be very informative—the range of scores is in fact very narrow across the sessions. In JCTE2, the mean accuracy score in trial 1 90.6%, and in trial 16 it is 92.4%. This suggests that dyads were performing close to ceiling throughout, and that they did not lose interest as the session progressed; indeed, if anything, they were trying to increase their score, as evidenced by the dyad mentioned above that deliberately discarded a slightly imperfect figure late in the session.

It is worth considering what exactly ‘accuracy’ is, in this game. Accuracy is not an aggregate property of the component pieces. It is a measure of how faithfully the component pieces have been put together relative to one another, and relative to the target. Accuracy is, in fact, a relational property. And we can also say something about what causes deviation from perfect accuracy. The reason scores do not reach 100% is presumably because, in assembling the figure, some of the component pieces have to be rotated before they can be put in place. Slight deviations from the target angle will cause slight discrepancies from the target figure, which can be seen in small gaps, of a

few pixels, between component shapes. Such discrepancies are probably impossible to avoid, since our visual systems are not set up to measure things in pixels, and neither is our mouse-control trained for such fine-grained work. Another potential source of discrepancy arises whenever judgement is required in where, exactly, two shapes should be stuck together. For instance, in Fig. 7.6, the pink square has to be glued onto the triangle above it somewhere along its bottom edge, but there is no definite point to anchor it to. By contrast, other target figures (such as the target assemblage in Fig. 7.1) can in principle be reconstructed simply by attaching pieces together at the corners—by attaching the corner of one piece to the corner of another piece.

This, I think, provides an important insight into what the task is, psychologically, from the point of view of someone carrying it out. The task is not about placing individual component parts in the position that is appropriate to *that particular part*. Instead, it is about attending to relations between parts. And this includes the gaps: the gaps are part of the task, and not merely part of the the background space in which the task takes place.

If the task is about constructing relations, then this would appear to problematize the use of this game as a tool for investigating the use of referring expressions (Bard et al., 2014). Or rather, it casts doubt on the assumption that what is being referred to is the set of component pieces. In fact, there are plenty of instances in the transcripts of things that look like referring expressions, but which do not refer to any particular shape. A few examples will suffice here, taken from dyad 7 in JCTE3. This dyad were quite economical in their use of spoken words. The following utterances appear to be actions aimed at coordinating a whole series of joins for the entire figure. They are uttered at the point in the trial after the pair have arranged all of the component parts separately into roughly the right places on the screen. From trial 7: ‘Shall we start from the top down?’. Trial 8: ‘Bottom up? Shall I start with the pink?’. Then in trial 11, the first utterance is, ‘Two big pieces?’ (another expression that does not refer to any one part). This is followed by: ‘Start at the top and come down?’ There are few ‘referring expressions’ here, of the type typically studied, yet these utterances are perfectly sensible. There’s no real need to refer to the individual parts: the parts themselves, as well as the relations between the parts, are directly visible to each subject and there’s no need to refer to them verbally. What the verbal actions are doing here is they are

coordinating activity. To assume that the referring expressions that are used must refer to elements, or objects, is perhaps to distract oneself from the true nature of the task, or to impose onto it a less-than-ideal description scheme.

Apart from continuing to be engaged with assembling the figures accurately, what else are the subjects doing? One possibility is that subjects may be taking on a role for themselves as ‘good subjects’; that is, they recognize that they have volunteered to participate in the study with the purpose of producing useful data for the experimenter, and they actively seek to produce useful data (Weber and Cook, 1972). There are some utterances in the corpus that appear to be directed as much at the experimenter (via the recording equipment) as at the partner. For instance, dyad 25 in JCTE2 engage in the following exchange, referring to the gaze cursor: ‘I find that blue thing really distracting actually... don’t know about you.’ / ‘um... the... eye th—... yeah I know’. The first utterance here sounds as though it may have been directed at the partner only in the second part after the ellipsis—as an afterthought. We might worry, then, that subjects are not *really* cooperating with one another per se, but are primarily cooperating with the experimenter, via the materials. However, the game is set up in such a way that subjects do in fact have to directly coordinate with their partner throughout: there’s no way to assemble the figures except by both members of a dyad engaging with the materials at the same time. We probably do not need to worry too much about good subject effects here.

We can therefore make the following conclusions about what the subjects are doing in this game. In general, they are genuinely are trying to be accurate. Probably secondarily, they are trying to be quick, or at at ant rate they tend to get quicker as the session goes on. They are collaborating with one another. And what they are collaborating on is a project of reassembling a target relational structure in as faithful a manner as possible.

7.4 Perseverative wiggling

It was suggested above that perturbation effects emerge in the JCT as a consequence of the way that the shuffling of conditions disrupts the subjects’ already-established action patterns. This description is quite vague and mysterious. To try to clarify what

is meant by this, it will be useful to look at a particular instance of an action-pattern that comes to be established during an experimental session. We will here focus on the phenomenon of shape-wiggling.

When a subject's mouse cursor is not projected on the partner's screen, it becomes more difficult for the partner to track what the first subject is doing. Not being able to see my mouse cursor means that you cannot be sure whether I am about to click on the same shape as you (and thus whether we as a pair are about to break that piece). And, moreover, it makes it difficult to coordinate a join. In order to know which piece to pick up, I need to know which piece *you* are holding. We both have to be holding complementary pieces for the join to be carried off successfully. One way to initiate a join, in the absence of the mouse cursor information, is for you to pick up some piece and to wiggle it about on the screen. I cannot see your mouse cursor, but I can see the piece moving. You then only need to wait for me to pick up a suitable second piece and to join it on to the first. This appears to be a standard solution that dyads adopt when completing a trial without the mouse cursor.

This wiggling action looks like a communicative signal. And of course it is; or, at least, that is a perfectly reasonable way to describe what is going on from the perspective of a third-party observer. For a subject engaged in the task, however, it may not necessarily be the case that the action is understood in these terms. Instead, it may be that in picking up and wiggling the first shape, you were simply acting on the scene in the most convenient manner that presented itself, as you attempted to join that piece to something else. Initiating a join might be conceived as acting on the relation between one's partner and the two pieces to be joined. In a with-mouse-cursor trial, this action can be carried out simply by clicking on one of the pieces (and perhaps waving the mouse cursor over the two pieces to be joined, or uttering some appropriate orienting expression). In a no-mouse-cursor trial, however, your simply clicking on the shape is insufficient as a means of acting on *me* (because I cannot see where you have clicked). You are forced to attend to relational structure of a higher order—to structure that specifies what *I* can see, rather than simply what *you* can see; the two are no longer the same thing.

The suggestion is that once you have performed this shape-wiggling action a few times, an action-pattern has been established. Producing a similar action in the future

may involve less effort, because you have already learned to attend to this particular type of structure in a particular way.

All of this is to introduce a phenomenon that is observed in some of the later trials, when the mouse cursor has been re-introduced. It is here no longer necessary to wiggle a shape as a means of coordinating a join. Subjects know this: at the beginning of the trial they are shown a message which indicates whether or not their mouse is being projected. Nevertheless, it appears to be quite common for subjects to continue to wiggle the shapes even though this is now redundant.¹ In other words, these subjects are perseverating on an action they have performed previously, much like the infants observed in the A-not-B studies.

The fact that subjects sometimes continue to perform this action when it is no longer necessary suggests that the ‘communicative’ interpretation of this wiggling behaviour is at best incomplete. Shape-wiggling is the kind of thing that is frequently labelled as ‘audience design’ (see chapter 5): it appears to require the ‘taking’ of another’s perspective immediately prior to the performing of the action. The existence of perseverative wiggling, however, suggests that this need not be the case. Here the subject is producing an action that appears to be ‘designed’, but not for the ‘audience’ in its present state. A dynamic systems interpretation might posit that the subject’s behaviour has simply been channeled by a pre-established attractor state.

An alternative explanation for this apparently-redundant wiggling behaviour is that it is in fact performed deliberately. Perhaps an action-pattern that is established during the no-mouse-cursor trials simply becomes part of an individual’s action repertoire. This action is subsequently called upon in later trials whenever it seems appropriate. This may well be the case, at least in some instances of late wiggling. It is, however, not possible to make definitive distinctions between deliberate and accidental instances from the video data.

Of course, it is important to try not to read too much into all of this. The present discussion is informed only by a handful of instances observed in the videos. But we can make at least one prediction about shape-wiggling based on the relational field model. We can predict that shape-wiggling is *caused* by the removal of the mouse-

¹Some examples, with approximate time points in the videos: JCTE3, dyad 6, trial 14 (29:00), trial 16 (34:30); dyad 7, trial 9 (14:20).

cursor. There is no need to wiggle shapes about if the other person's mouse cursor is visible. There should therefore be little shape-wiggling in trials in the first block in which dyads could see each other's mouse cursor.

The important idea here is that the establishing of an action-pattern is costly in terms of effort—it requires exploration of the relational field. Subsequent instances of that action-pattern, by contrast, are cheap: a pre-established behaviour can be invoked rapidly or even automatically. To return to the overall pattern of perturbation across the session, the suggestion is that the costs (in terms of slower trial completion times, and so on) are a consequence of action-patterns no longer being appropriate to the circumstances—to the resource layout. The gains, meanwhile, are an index of increased adaptedness of the subjects' behaviour patterns as the session progresses.

7.5 Performing by looking

The JCT was set up as an eye-tracking study. Specifically, it is designed as a tool for studying eye movements during a collaborative activity (Carletta et al., 2010). The corpus therefore includes a vast amount of looking data, which means it is possible to examine subjects' looking behaviour at a quite basic level, in terms of where, specifically, subjects are looking, at any given point in time. This data we cannot as easily treat in terms of perturbation, but we can again get some ideas about how people make use of the available visual structure, which will allow us to make further predictions about what would happen if this structure *were* to be removed.

Some eye-tracking studies on individual subjects have investigated people's eye movements during the carrying out of everyday, routine tasks, such as making a cup of tea or preparing a peanut butter and jelly sandwich (Land et al., 1999; Hayhoe, 2000). These kinds of tasks are well-learned, to the extent that introspection tells us that we perform them without really having to pay much attention to what we are doing. The eye-tracking data, however, reveal that, during the task, the eyes fixate almost exclusively on the objects that are relevant to the task as a whole, and do not wander off to examine some other part of the visual scene (even though one might expect that 'mind wandering' should be accompanied by 'eye wandering'). And, moreover, the particular objects that are fixated at a given time are the objects that are currently being

acted upon, or are about to be acted upon. In a paper summarizing these findings, [Land and Hayhoe \(2001\)](#) conclude: ‘our studies lend no support to the idea that the visual system builds up a detailed model of the surroundings and operates from that. Most information is obtained from the scene as it is needed.’

If this is so for individual tasks, it might well be true of the JCT as well. We can get a rough idea about where subjects look during a JCT trial by examining the graphs in Fig. 7.7. The corpus is coded into distinct temporal phases. The time that a dyad spends joining two figures together is defined as a ‘dock’ phase. Everything that precedes a given dock phase is further divided into two equal-length sections. The first of these is defined as a ‘plan’ phase, and the second as a ‘process’ phase.² Meanwhile, looking-direction is coded spatially, in terms of the specific screen region being fixated at a given point in time. The screen is divided into several regions, the most important of which are the ‘target figure area’, in the top-right of the screen; the ‘work area’, where the pair’s figure is assembled; and the ‘initial parts area’, underneath the target figure, where component parts are found at the beginning of a trial.

The graphs in Fig. 7.7 display the proportion of the total count of ‘looks’ that are directed to a specified screen region.³ The graphs in these figures are qualitatively very similar to one another. One feature that is common to all three graphs is that looks to the target figure occur most frequently in the ‘docking’ or joining phase. This effect is particularly pronounced in JCTEs 2 and 3. Clearly, this somewhat contradicts the phase labels themselves: if the first stage were really a ‘planning’ phase, then one would expect that this is when the target figure should be fixated most. Instead, the data appear to show that the things people look at are the resources that are relevant to the current section of the task, much as in [Land and Hayhoe’s](#) food-preparation studies. It does not appear that subjects are memorizing the layout of the target figure and then building their own figure based on that internal model. Rather, they are building the model piece-by-piece, making use of perceptual relations between the assembled figure and the target as they become relevant.

²Two other types of phase are also defined in the corpus, but not considered here: one for the time spent recovering from a break, and the other for the wrapping-up section at the end of a trial.

³Note that this is not the same as the proportion of overall looking *time*. A ‘look’ is defined here as a series of fixations within a single region; the look ends when the eye fixates something outside that region.

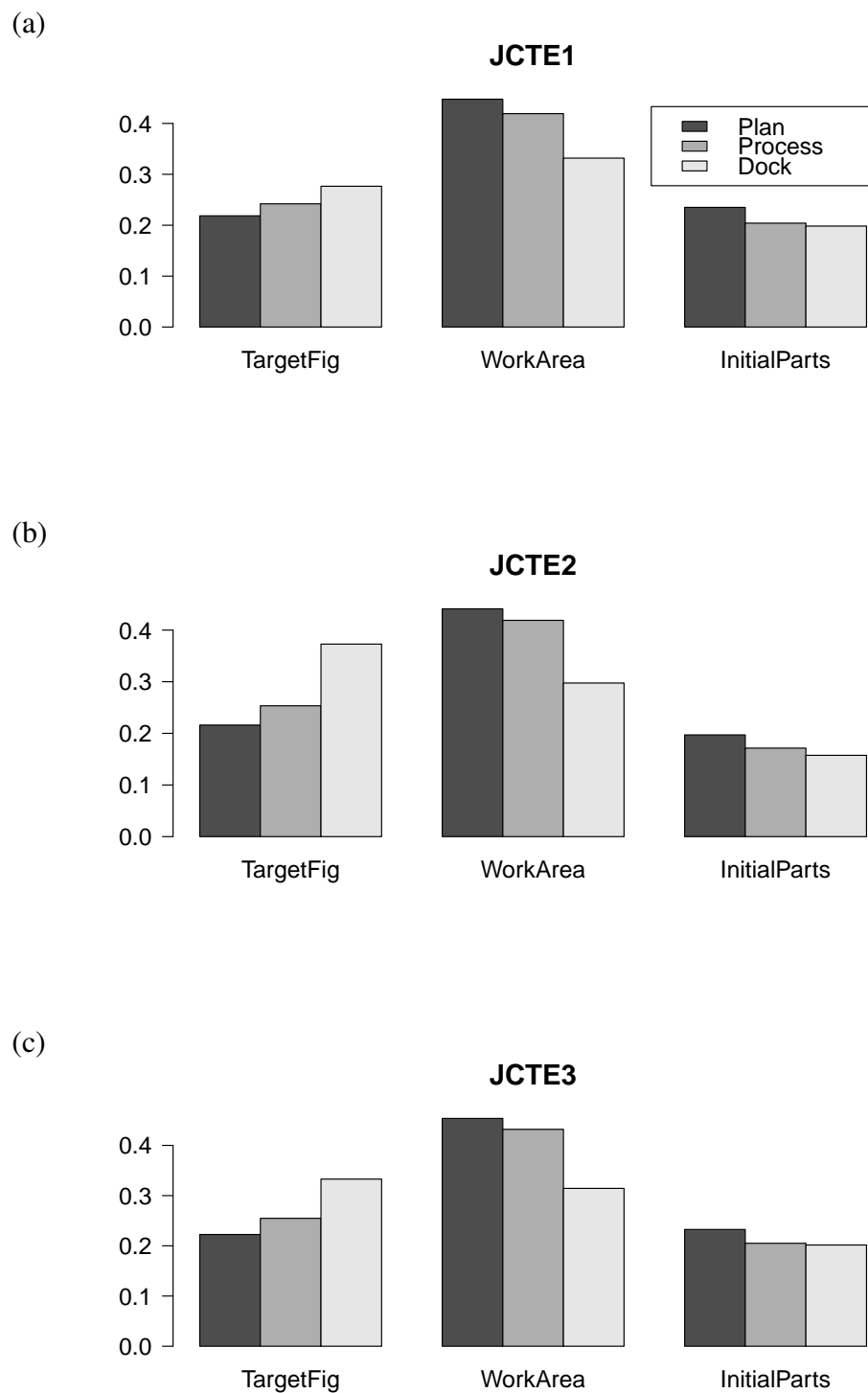


Figure 7.7: Proportion of looks to specific screen regions during a join ('dock'), and during the temporal phase preceding a join, which is split into two halves (a 'planning' phase and 'process' phase), in a) JCTE1, b) JCTE2, and c) JCTE3

This looking behaviour is worth noting here because it brings into doubt an assumption that has been mentioned previously: that performing a task collaboratively requires the sharing of mental content, or the building up of ‘common knowledge’. Subjects in the JCT, just as in the individual tasks, appear to be using structure in the environment as a set of resources that are reliably present throughout the task. The environment is performing at least some of the work that might otherwise be attributed to a posited store of shared content.

And we can make the following prediction: perturbing visual access to this structure, say by making the target figure visible only at the start of a trial, should reduce the dyad’s performance on various measures, including accuracy. Of course, this is not a very interesting prediction. A traditional cognitivist could make the same prediction, appealing to working memory capacity instead of perturbed dynamics. A more interesting prediction, then, is that we’d also see a perturbation effect if visual access to the target figure were denied only at a strategically important moment—in the ‘docking’ phase. One way to implement this would be to mask the target figure whenever both subjects have moused-down on (two different) component parts. This set-up would prevent them from consulting the target figure—and, more specifically, the relations between the target figure and the component pieces—at precisely the moment when this information appears to be needed. We can predict that this will affect performance, and that there will be some cost in adjusting to it (in terms of temporarily increased trial durations, more mouse actions, and so on).

7.6 Some open questions

Clearly, the JCT data set is quite rich and it could potentially be explored in a large variety of ways. It would not do, however, to lose sight of the fact that the shape-assembly task here is an artificial one and one that is therefore not inherently informative about real-world behaviour. To address the data only in terms of questions that are internal to an understanding of this particular task would be to treat the data set as a tool for generating research for its own sake. It will perhaps be more productive to use the data as a means of generating further questions to explore, outside of this data set itself. I will here briefly discuss two such questions.

The first question that arises concerns the potential uses of the gaze cursor. It was suggested above that the kind of event structure that is produced by this cursor is structure that subjects are likely to find difficult to use, at least within a single experimental session in which they have other activities to complete. But can a gaze cursor ever be a useful source of information? The answer to that is that, yes, it can be. A study by [Brennan et al. \(2008\)](#) presented pairs of subjects with a quite different activity to complete. Here, the task was to find a letter ‘O’ hidden in a screen filled with ‘Q’s. Again, two subjects were hooked up to eye-tracking devices, and each subject’s gaze cursor was projected on their partner’s screen. In contrast to the JCT, however, the pair were here seated in different rooms. In some trials they were also not allowed to speak to one another, and the gaze cursor was therefore the only available source of information about what the other person was doing. The finding from [Brennan et al.’s](#) study was that not only was gaze information useful, but interpersonal coordination was in fact most efficient in the gaze-only condition (with no speaking). Of course, the coordination task was quite simple: the easiest way to divide the labour in the visual search is for each subject to focus their own search behaviour on the side of the screen where their partner is not looking. And here is the important difference between gaze information in the visual search task and in the JCT. In the visual search task, the gaze cursor is inherently meaningful in the task context, because the task itself is about searching a visual space. By contrast, in the JCT, looking behaviour is at best only indirectly meaningful. In the JCT, the real business is conducted in the currency of shape movements, rotations, and mouse clicks. The open question is this: could a gaze cursor be useful as a tool in coordinating behaviour in more complex activities than the one devised by [Brennan et al.](#), or is it only good for coordinating behaviour in visual search tasks? And in either case, what applications might this gaze-cursor have outside the laboratory?

The second question concerns the relationship between perturbation and learning: does perturbation merely impede learning, or can it drive learning in a useful way?

It has often been observed that interaction appears to enable the emergence of ‘abstract’ solutions to specific problems. For example, dyads outperform individuals in finding efficient mathematical solutions to practical problems, such as how to multiply fractions ([Shirouzu et al., 2002](#); [Schwartz, 1995](#)). And a finding from a number of

referential communication games is that ‘abstract’ referring systems emerge reliably in the course of the interaction. In the maze task, in which one subject describes a route to another (Garrod and Anderson, 1987), dyads start off by describing salient features of their mazes, and move on to referring to individual boxes in the maze using a coordinate scheme (‘a, 1’). One suggestion is that such abstraction results from misunderstandings that occur during an interaction (Healey, 2008). A problem with this is that it is not clear that either of these words—‘abstraction’ and ‘misunderstanding’—has any kind of precise meaning. If these words are not being used precisely, then they cannot be doing any explanatory work; they are merely being used to describe a phenomenon of interest.

The relational field model can potentially provide a more precise set of terms here. The things we are inclined to label as ‘abstract’ simply involve attending to and exploiting higher-order relational structure. And conversely, anything that we are inclined to think of as ‘non-abstract’ (i.e., ‘figurative’ or ‘iconic’) involves attending to only more basic relational structure: to objects and to first-order relations between objects, but not more. Misunderstandings, meanwhile, need not be things that individuals actively attempt to detect (as suggested by Healey, 2008). Instead, misunderstandings arise as barriers to the carrying out of a particular, local action that an individual is attempting. A misunderstanding acts as a kind of perturbation of the actor as a system. It specifies, to the actor, that the current action being attempted is not suitable for the purpose it is intended for. In other words, it provides information to the actor that some sort of behavioural adjustment is necessary.

There is a particular type of ‘abstraction’ that arises in the JCT, as we have seen: removing the mouse cursor means that subjects have to compensate by attending to higher-order relational structure that specifies a difference between the two screens. One way that they can adjust is by wiggling shapes around on the screen. The wiggling behaviour appears to emerge as a consequence of a particular type of perturbation, and the suggestion has been that this learning to wiggle is a *consequence* of the introduction of a particular type of difficulty.

If this is true, it is relevant to an idea from educational psychology which says that learning is more effective in the long-term if ‘desirable difficulty’ is introduced into the teaching materials (Bjork and Bjork, 2011): the materials are structured so as to require

the learner to ‘actively’ seek meaning, rather than ‘passively’ observe it; for example, by making the overall structure less clear. This phrase—‘desirable difficulty’—is another that seems intuitively appealing but that lacks precision.⁴ And again, perhaps this idea can be made more definite within the framework of the relational field model. The basic idea is that learning is more effective, or at least more ‘abstract’, when it involves perturbation followed by adjustment, and followed by attending to higher-order structure. The question is: can this way of thinking provide a practical way to judge which types of difficulty are desirable, and which are merely difficult?

7.7 Summary

From this brief examination of the JCT, we have seen that the relational field model, enhanced with some ideas from dynamic systems theorizing, can serve as a tool for analysing a human collaborative activity in a potentially novel way. Here, structure in the environment is seen as part of an entire system which also encompasses the actor. We have seen that behavioural perturbations occur when the environment changes. In the JCT these environmental changes are inserted artificially, but elsewhere they may arise more organically, as people learn to carry out a task. It appears that some of these perturbations can lead to novel behavioural strategies (as in the adoption of wiggling following removal of the mouse cursor).

A further finding comes from the eye-tracking data, which suggest that, in carrying out the task, people make use of persisting structure in the environment; subjects look at this structure exactly when it is needed. External structure is instrumental in carrying out the task. This observation further erodes the mystery of ‘common knowledge’: there is no need to appeal to shared mental content here, because what is shared is what exists externally: structure in the environment that is readily available for either subject to perceive. These ways of looking at the data reveal the importance of agent–environment reciprocity, which, of course, is exactly what the relational field model is designed to address.

⁴But note the similarity to [Vygotsky’s \(1978\)](#) notion of the zone of proximal development, as discussed in [chapter 2](#).

Part III

Closing remarks

Chapter 8

Lessons learned and prospects for future research

I hope to have demonstrated in the course of the preceding chapters that it is possible to study social phenomena on ecological realist principles. Actions carried out in a populated environment—speaking, collaborating, competing—are conceived as varieties of acting on a field of relations. This puts activities carried out with others on the same footing as activities carried out by an individual acting alone. An individual acting upon an inanimate object or exploring a space is also acting on the relational field. In everything it does, the animal occupies a place in a complexly structured field of relations, and all actions, social or otherwise, are actions on that field.

This conception of acting with others will prove valuable if 1) it can be further built into a more complete theoretical approach while maintaining its coherence, and 2) it can be usefully applied to the solving of practical problems. I will use the second half of this final chapter to outline some promising directions in which I believe the model can be developed. In terms of theory, I suggest that the relational field model offers a useful framework for a general psychology of learning: learning is not conceived as something that happens exclusively or primarily in childhood, but as a lifelong process of change that occurs as an individual explores and negotiates the structure of its environment. In practical terms, meanwhile, the model may be able to generate a useful set of tools that can be applied in various areas of design, such as the design of public spaces.

First, though, it will be worthwhile to try to bring together the main points that have been raised in the discussion so far. I will not repeat the entire argument that has been made. A brief summary has been provided at the end of the key chapters; I will avoid recapitulating that material here. Instead I propose here a list of the lessons that I have learned in writing the thesis—lessons that have been left implicit in the text, or else have been mentioned in passing even though they constitute non-trivial ideas that deserve to be scrutinized in their own right. This list is offered in the spirit of Jacob Cohen's (1990) observations on the use of statistics in psychology, which he delivered in an address to the American Psychological Association under the title 'Things I have learned (so far)'.

8.1 Things I have learned about the psychology of speaking and collaborating

Just because we give something a name, doesn't mean it's real.

Psychology as a discipline finds itself in a continuous struggle with its own terminology. There are two major sources of troublesome terms. One is everyday speech, the other is the set of metaphors that psychologists themselves have developed to describe things. We will consider everyday speech first. The basic problem arises from the fact that the set of technical concepts that psychologists have constructed overlaps to a great extent with the set of words we use in everyday practice for describing the behaviour of others (Ryle, 1949). There is nothing inherently problematic about using plain, everyday terms in a technical context. The problem is that some of the words that are borrowed are unsuitable for the purposes that they are put to. In particular, everyday speech makes frequent and casual use of *nouns* to characterize behavioural phenomena—we speak of people as having 'a good memory', of engaging in 'deep thought', of having a good command of 'language', and so on. But a noun in everyday speech need not necessarily correspond to any particular component of the actual machinery underlying the behaviour being described. Behaviouristically-inclined psychologists in the first half of twentieth century recognized this fact, and scrupulously avoided nouns in favour of verbs. Verbs function to delineate a phenomenon of inter-

est, but without implying a particular way in which the underlying machinery must be arranged. Thus, instead of describing people as possessing a ‘memory’, it is said that we engage in a process of ‘remembering’; instead of containing ‘thoughts’, people are said to engage in ‘thinking’ (Bartlett, 1932, 1958). This remains a useful technique for guarding against the reification of constructs.

Some nouns to be suspicious of in the context of interpersonal behaviour include: ‘group’, ‘common knowledge’, ‘common ground’, ‘collective intention’, ‘belief’ (and ‘false belief’), ‘role’, and the various synonyms for ‘contagion’ (see the glossary for references to discussion in earlier parts of the text).

The second source of potentially misleading names is the set of metaphors that are used to make sense of psychological findings. A common approach in psychology has been to take the individual animal as the unit of analysis, and to model that animal by analogy with some particular piece of technology; often the most advanced piece of technology that psychologists feel they are familiar with—telephone exchanges, digital computers, statistical software packages, and so on (Gigerenzer, 1991; Barrett, 2011). The names of the parts that make up the particular technological artifact in question can then simply be borrowed and used as names for the hypothetical parts of the animal’s psychological machinery. Everybody knows that these names-for-parts are to be used strictly metaphorically. Still, a theory that is built around a metaphor inevitably comes to be shaped by that metaphor in important ways. An example is the idea that the mind is a slate (blank or otherwise), i.e. that the mind is a device for storing symbols. In chapter 1 it was suggested that the slate metaphor encourages a conceptualization of learning as passive accretion, and distracts attention from the dynamic nature of learning as a thing that an animal does. Learning might instead be characterized as change through active exploration. The doctrine of the new ball was proposed as an alternative metaphor—one that encourages a view of learning as change. It is important that we choose our metaphors carefully, and that we occasionally pause to examine those already in widespread use: we must be wary of the influence of borrowed names.

The fact that we don't have a name for something does not mean it is fictional.

The ecological realist approach does not seek to model the individual per se, but seeks instead to explain the workings of a broader animal-environment system. The machinery of this system requires its own set of names: energy arrays, invariant structure, surfaces, transformations, and so on. It has been argued that the animal-environment system consists of a set of physical entities held together by a complex network of relations. A difficulty is that relations are not things that we typically give names to (at least not in the English language, which is the language employed in nearly all academic discourse). We describe one object as being bigger than another, but we do not talk of an object having 'bigness' over anything. There are words that could be thought of as names for relations: 'congruity', 'mismatch', 'discrepancy', 'equality', 'ambiguity', 'disdain'. But we tend to think of these as 'abstract concepts'—as things that are not really out there in the world, but are merely a way of describing and comparing things we have perceived: they are thought of as ways of comparing mental objects or percepts. The word 'affordance' is a name for a type of relation; it is a name that had to be invented (Gibson, 1966, 1979).¹ Relational thinking is hard to sustain simply because our natural language does not recognize the existence of relations as perceivable entities. The present thesis has been concerned with applying ecological realist principles to an understanding of interpersonal phenomena such as speaking and collaborating. The claim is that in order to make this project work it must be recognized that relations are real—that they are a part of the world that is perceived and acted upon and not merely a property of a mental description of that world. Relations constitute a crucial part of the explanatory machinery of an ecological approach to social activity. This is true whether or not it accords with our everyday descriptions. Clearly, it will be necessary to invent some novel technical terminology: 'relational fields', 'relational webs', 'calibration'. This terminology will have to be developed carefully, and with due attention given to the fact that simply giving a name to a phenomenon is not the same as explaining that phenomenon (Dunnette, 1966).²

¹Again, it is by no means universally agreed-upon that 'affordance' does label a relation (Chemero, 2009); I assume for present purposes that it does.

²It is also not the case that new terms can straightforwardly be invented from whole cloth, or imported from other fields. If we are to make sense of them, the terms we use must be rooted somehow in existing

Don't be afraid to treat people like animals.

In the course of the thesis, I have attempted to draw on a broad range of examples of collective activity. In particular I have been careful to consider activities carried out by groups of non-human animals: pack-hunting in chimpanzees and wolves, and the alarm calling behaviour of vervet monkeys and other species. These are fascinating phenomena in their own right. They also serve as demonstrations that collective activity need not necessarily depend on the workings of complex linguistic apparatus in the head. We are not inclined to attribute elaborate propositional mental states to animals. This is a healthy attitude: it encourages us to search for causal structure in the animal's environment, and discourages us from jumping to the conclusion that the important business of coordinating action must be occurring in a not-directly-observable inner world (for further examples of collective activity in ants and robots, see [Barrett, 2011](#)). The same methodological approach can profitably be applied to human collective activity.

Certainly, any approach to collaborative activity that cannot accommodate the collective behaviours of animals is an approach that is limited in an important way. Such an approach has probably over-intellectualized its answer to the question of what it is to collaborate. It appears to be commonly assumed that in order to engage in some activity with others, the individual must be in possession of some internal propositional statement which represents to the individual the fact that this is indeed a collaborative situation. It should be stressed that to question that collaborating always involves manipulating such internal content is not to deny that language ever has a role to play. Clearly a great many of the collective activities engaged in by humans involve the use of verbal actions. But it's easy to over-emphasize the role of language in coordinating the activity as a whole. One way to guard against this is to develop an approach that deals equally with animal and human collective activities. The present approach says that collective activities are activities carried out in a relational field. Verbal actions, like physical actions, are simply actions on that field. Viewing speaking in this way has the further advantage of eliminating language-environment dualism—the implicit as-

usage, in psychology or in everyday speech. [Cutting \(1982\)](#) states an evocative case against careless borrowings: 'borrowed terms are deracinated; they are cut off from the roots of their own disciplinary matrix. At worst they shrivel and die; at best they grow to be something different than they were.'

sumption that language is a special kind of substance that interacts with the substance of the environment but is not itself part of that environment (more on which below).

Don't confuse your own perspective with that of your subject matter.

I have already mentioned the problem of the psychologist's fallacy. This occurs when investigators mistake their own third-party description of some phenomenon for a description of the causal machinery underlying that same phenomenon. This problem is not unique to psychologists; a similar fallacy can be committed by investigators in any of the sciences concerned with animate beings. Call this the fallacy of the analyst's perspective. Thus, vervet monkeys are said to 'make excellent primatologists', on the grounds that they behave as if they recognize the social dominance hierarchy of their group, and as if they recognize others as out-group members, and so on ([Cheney and Seyfarth, 1985b](#)). But monkeys were engaging in these behaviours long before any primatologist came along and described them. If anything, primatologists should instead be concerned with whether or not they themselves are doing a good a job of being monkeys.

In developmental psychology, it sometimes appears to be standard practice to conflate the researcher's perspective and the subject matter: the metaphor of the child as scientist is often taken as a basic working assumption (for criticism, see [Kuhn, 1989, 1992](#); [Donaldson, 1978](#)). Children are said to be 'intuitive scientists'—Kuhnian problem-solvers rather than Popperian falsificationists ([Karmiloff-Smith, 1988](#)). What this metaphor presupposes is that the world is imperfectly perceived and that the child is engaged in a project of *trying* to learn about how things work: the child is trying to build up an internal model of the world. But there is an alternative view, which I have adopted, and which I alluded to above with the doctrine of the new ball. The alternative view says that children simply engage in activities, and by doing so they change *themselves*. The change happens gradually as a consequence of the child's practice and experience of acting and attempting to act ([Thelen and Smith, 1994](#)).

There are versions of the psychologist's fallacy that are specific to the study of collaborative phenomena. Commonly, the analyst attempts to explain what is going on at the dyad or group level. From the analyst's perspective it appears quite reasonable to describe the group as engaging in a joint activity, or as possessing common ground,

or as exhibiting desires or beliefs. But these labels refer to properties of the group as an entity that has been artificially delineated from its surroundings by a third party. There is no reason to assume that these labels correspond to any part of the causal machinery itself. The fallacy of the analyst's perspective, operating in research at the group level, has created some deep confusions: that there is such a thing as a genuine joint activity (which exists as a natural kind); that the group should be seen as a kind of agent (Allport, 1924); or that acting with others involves a special kind of 'shared intentionality'. The relational field model avoids these perspectival confusions. On this model, what the subject sees is what the analyst sees: a richly structured environment filled with relational structure. The problem, for the analyst, is to identify the particular components of this structure that are instrumental in the carrying out of some task under investigation.

Point your regresses outwards.

A major aim of the present thesis has been to outline an account of the psychological machinery that must be necessary to explain the fact that people are able to engage in collaborative activities. It is customary, when proposing such an account, to worry that the set of assumptions being put forth might in fact entail some sort of infinite regress that, after all, makes the entire proposal unworkable. Indeed, for a representational account, this is a real possibility. Suppose that I engage you in some collaborative activity: How do I know that you know what actions to carry out? And how do I know that you know that I know what I am doing? And so on, ad infinitum. The problem here is that an infinite set of such calculations presumably cannot be carried out by a single, finite brain. The regress is pointing inwards, and an inwards-facing regress is a runaway process. By contrast, the non-representational account put forward in the relational field model does not face this problem. Collaborative activity is coordinated, it is proposed, via relational structure that exists in a shared environment. This relational structure may itself be infinitely complex: there can be relations between two objects, relations between an object and a relation, between a relation and another relation, and so on. But this outward-facing regress places no inherent burden on the internal workings of a given animal. An outward-facing regress is instead a feature of the world that an animal can exploit. It is an example of the kind of convexity that is character-

istic of natural systems, and that makes such systems ‘antifragile’ (Taleb, 2012). A self-contained computer system is fragile in that it is liable to crash whenever it is set to work on an infinitely recurring calculation. Collaborative activity is the opposite: it is a way of exploiting the infinite ‘processing capacity’ of the relational structure of the world. Human language is also a tool that functions to act on this structure. Language appears to be a tool uniquely suited for exploiting an infinitely structured world that is filled with relations.

Make assumptions; but try to be aware of them.

Research on dialogue that is carried out in the tradition of communication theory maintains a strict language-environment dualism. As we have seen, this encourages researchers to focus their attention upon a certain kind of communicative phenomenon, such as communication games that can be carried out over a telephone line. The relational field model seeks to avoid this dualism, and as a consequence it must reject the metaphor of speaking as information-exchange or information-sharing. In rejecting this metaphor, however, the field model necessarily adopts new assumptions of its own: speaking is acting on relations; speakers and addressees must actively calibrate their relational webs to fit the present circumstances; and so on. Having a set of assumptions is no bad thing. On the contrary; without making assumptions at some level it would be impossible to make any kind of empirical claim about anything at all. An aim of the present thesis (as mentioned in the preface) has been to demonstrate that it is important to make explicit the assumptions that we *do* make. This is necessary if we wish to be able to properly evaluate the claims being made.

What we cannot claim is that one set of assumptions is necessarily better than another—that the relational field metaphor is inherently better than the information-sharing metaphor, say. There is no way to determine, a priori, which of these metatheories is most likely to lead to empirical success, or to practical outcomes. This fact alone is a persuasive reason to accept methodological pluralism (Chemero, 2009). Another reason to favour pluralism is that maintaining multiple metatheories can make any given theory stronger: partisans of a given theory must engage with competing programmes, and must work to accommodate those programmes’ findings. This last idea has been called the method of multiple working hypotheses (Platt, 1964). It is a

principle that appears to be seldom adopted by psychologists (Reed, 1996a).

Don't try to answer all the questions at once.

A good metatheory is one that generates research findings that build cumulatively, each finding allowing researchers to ask further questions, or to better make sense of what has already been established. However, in the study of dialogue and interpersonal activity, it has been common for effort to be directed at the project of building grand theories that try to explain everything at once. These theories seek to posit a handful of basic principles that can explain all possible facts about acting with others: collaborating is said to be at heart a form of mind-reading, or it is said to rely fundamentally on processes of interpersonal contagion, or whatever. The temptation, then, is to argue about which of these grand theories gets closest to the truth. Consequently, research effort is directed at attempts to construct some laboratory-based activity whose results will neatly adjudicate between the competing grand theories (notice that this project is in stark contrast to the spirit of the method of multiple working hypotheses). The concept of the task, as defined in chapter 4, is a response to this urge. The task is a methodological device that directs research attention towards specific, well-delineated problems, and away from intertheoretical disagreement. The ecological approach in psychology has proved successful precisely because it has taken well-defined tasks as its unit of study. The concept of the task enables a genuinely cumulative research programme. Another of the main arguments of the present thesis has been that this concept of the task-as-unit-of-analysis can usefully be applied to the study of activities involving multiple individuals—speaking, collaborating, competing; in short, all of the activities engaged in by animals who find themselves living in an environment populated by other actors. The concept of the task motivates a practical empirical programme. The previous two chapters have attempted to establish that the relational field model is of potential use for guiding research to be carried out within this programme. In the section below I will briefly outline some specific research areas where I believe the field model can be applied in a productive manner.

8.2 Prospects for future research

In the present thesis the relational field model has been developed specifically as a way of describing social phenomena while avoiding some untenable dualisms: between language and the environment, and between the environment and other actors. However, a major reason to adopt the relational field model may be that it provides a way of unifying the study of psychological activities that have come to be treated as separate phenomena: developmental, social, clinical, organizational, cognitive.

Of course, there is a reason why the different parts of psychology have come to be treated as separate, specialized subfields. The reason is that the individual was taken as the unit of analysis, and it was recognized that to study the entire individual all at once was too difficult a project. Research specialization arose, no doubt, as a natural consequence of different researchers pursuing seemingly different goals, and gradually losing contact with one another as their own subfields expanded to occupy the majority of their time and resources. A formal justification for this state of affairs was the suggestion that the mind is divided into modules, and that those modules can be studied in relative isolation from one another (Fodor, 1983). The modular view of the mind justifies a modular approach to research, and in turn it necessitates such a divided approach.

The modular view of the mind does not, however, appear to provide a suitable foundation for the study of collaborative activities. A collaborative activity does not occur in the brain, or at least, it does not occur in any single brain, but is distributed between multiple individuals and across space and time. Collaborative activity demands a broader conception of the subject matter of psychology: what is to be studied is the set of events that occur in an animal-environment ecosystem.

The ecosystems view of acting has proven useful for studying fairly simple perception-action phenomena in individuals (Warren, 2006; Wilson and Golonka, 2013). The relational field model encompasses these activities of individuals and expands the range of enquiry to include social activities as well. I have restricted myself to discussing activities carried out by small groups of mostly co-present actors. But there is no reason in principle why this ecosystems perspective cannot also be applied to phenomena that are the traditional concern of sociologists and economists. In economics, as in psychol-

ogy, standard practice privileges the individual as the unit of analysis. The assumption is that individuals are rational maximizers of their own utility—that is, it is assumed that people are driven by a desire to accumulate personal wealth. In reaction to this, a number of economists have advocated precisely an ecosystems view of economic activity. This is motivated by the conviction that the standard individualistic view has proven destructive at the level of the wider environment (Costanza, 1989; Arrow et al., 1995). The relational field model provides a plausible psychological foundation for an economics concerned less with individual accumulation and more with maintaining the broader environmental system of which the individual is a part. On the individualistic view, a given transaction is understood in terms of personal costs and benefits for the parties immediately involved. Environmental consequences are referred to as ‘externalities’ (which appears, in effect, to mean ‘somebody else’s problem’). On the ecosystems view, the environmental costs cannot simply be ignored, but have to be recognized as an inevitable part of a specific way of going about business. Transactions do not occur in some realm separate from the environment, but take place *in* the environment. Harmful, polluting effects are therefore things that an economic agent chooses to inflict on the environment in choosing to carry out a particular transaction, and are to be recognized as such: ‘externalities’ are things that economic agents are to be taxed and held accountable for.

I mention the above to illustrate that the relational field model is flexible and can provide a coherent framework for studying a range of phenomena that are typically studied in isolation from one another, and are compartmentalized into different research fields: psychology, economics, human geography. I will here focus on one particular question that is internal to psychology: if the relational field model is adopted, what implications does that have for the status of the psychology of learning?

8.2.1 Learning

Learning has a curious status in modern psychology. It is given a high degree of prominence and is the subject of a substantial amount of research. Yet it remains somehow separate from the kind of experimental psychology that typically studies adults. A clear expression of this arises in Cheney and Seyfarth’s (1990) comment that children

(and vervet monkeys) are to be studied as ‘almost minds’. Learning is thus seen as something that children have to do in order to become fully-fledged actors. But it is not considered an important part of the activity of adults carrying out routine tasks.³

One area where adult learning is taken into consideration is in the study of the acquisition of expert performance. The particular skills that are traditionally studied here are chess, sporting activities, and playing a musical instrument (Ericsson and Charness, 1994; Ericsson et al., 1993). That these are the activities that are studied perhaps gives the impression that adult learning is a luxury that adults can engage in if they happen to have some spare time on their hands. But the study of the acquisition of adult expertise appears to be based on an important assumption which is also made in the study of child learning. In both of these areas, there is an implication that learning is something that has to happen first, and only after this learning has been completed can performance proper begin. Thus, Ericsson et al. (1993) set out to explain expert performance ‘as the end result of individuals’ prolonged efforts to improve performance while negotiating motivational and external constraints.’

In ecological psychology, too, learning is typically studied separately from the basic study of perception-action tasks. There is a historical reason for this. In the early 1960s James and Eleanor Gibson decided to split their research activities, respectively into the study of the senses and the study of perceptual learning (Gibson, 1966, p. viii). This appears to have worked quite successfully: James Gibson’s account of the structure that exists in the environment to be perceived fed in to his wife’s work on how infants learn to discriminate that structure (Gibson and Pick, 2000). However, in the cases of acting collaboratively and speaking, this split is less sustainable. Actions that are carried out with others are not simply out there in the environment to be perceived and explored, like staircases and puddles are. Rather, collaborative actions are created in the very act of exploration. On the account of language learning outlined in chapter 3, learning to speak relies on the child’s attempting to speak. If this is correct, then to understand what speaking is, it is not sufficient to characterize the ‘input’ (the ambient verbal activity that a child is exposed to); it must be recognized that the child is partly responsible for creating that input. The special character

³Actually, in many laboratory studies carried out with adult subjects, learning is taken into consideration, but it is considered as a confound to be controlled for: a ‘learning effect’.

of learning to collaborate with others is that the objects one learns to collaborate with are themselves animate perceiver-actors who themselves learn—that is, other people re-calibrate their own personal relational web in response to actions carried out by the child. A populated environment is an environment that acts back.

On the account that has been outlined, learning is not something that happens before performance can begin. From the first-person perspective of a perceiver-actor, there is of course no point-of-transition that is reached at which the learning phase ends and the performance stage begins. Rather, learning is a lifelong process of change that simply happens as the animal engages with its environment, and as it attempts repeatedly to manage that environment for its own needs. A similar conclusion was reached by E. B. Holt, one of the original new realists. Holt further argued that a person's *ethical* conduct towards others emerges as a consequence of the individual's ongoing engagement with the facts presented by the populated environment, and of their evolving capacity to discriminate those facts (1915, p.197): 'This matter of the unthwarted lifelong progress of behavior integration [learning] is of profound importance, for it is the transition from behavior to conduct, and to *moral conduct*.' On present terms, the learning that occurs across childhood or over extended periods of practice at some activity *is simply the same thing* as an individual's calibrating to its immediate surroundings. The difference is just that these are observed at different scales. Learning is calibration observed over a longer sample of time.

And if learning in general is lifelong, it will be necessary to think of language-learning, in particular, as something that never stops, during the lifespan of an individual. Chapter 5, which discussed the structure of referential communication games, concluded that speaking in interaction is a constant process of calibrating one's relational web to local circumstances. Chapter 7 used an analysis of behaviour in the JCT to make the case that learning to speak in a particular, novel way is a natural part of learning to engage in a novel activity. If we view this kind of calibration to local circumstances as a kind of language learning, then it is not the case that there's any kind of essential separation between the behaviour of children and adults—with children learning a language and adults merely using a language that has already been acquired. Instead, there is a continuous process of developmental change. This process of change does, however, make adult behaviour more flexible than child behaviour: the

adult is able to attend to more complex forms of relational structure as a result of past experience. As a result, the calibration process becomes increasingly efficient as the individual develops. It is, however, not the case that because adult learning is efficient it can be taken for granted. Chapter 7 aimed to show that when a learning perspective is applied at the level of a single experimental session, in which a game is carried out by adults, it is possible to see learning taking place as the session unfolds. This kind of analysis allows us to investigate how the construction of an environment interacts with the dispositions of actors and allows us to identify the specific environmental resources that are instrumental in actors' ability to complete the tasks they wish to complete.

Even if we accept that language learning never stops, however, we can still maintain a particular interest in how children learn to speak at the earliest stages. If we adopt the relational field perspective on infant language learning, a number of questions follow. Is it the case, as was suggested in chapter 3, that children have to learn that they can produce the same sound in the presence of different adults and it will have the same meaning? If so, how is this achieved? And how do children in bilingual settings learn to differentiate their verbal actions into two types? Is the account outlined chapter 3 a useful one? And is it possible to intervene in the child's learning of relational structure? Is there some manipulation that can be performed to encourage children to explore, say, second-order relations, at an earlier stage than is typical?⁴ To fully address these questions it will be necessary to develop novel experimental techniques. But of course there already exists a great deal of published data about children's language learning, and there are many corpora of children's early vocal productions. Some of this data will be relevant to these questions. Is there, for instance, corpus evidence that young infants produce specific sounds only in the presence of specific caregivers? There is much to investigate here.

⁴A neat intervention paradigm has been used to investigate how infants learn to interact with objects; that is, with first-order relations. In these studies, a pair of 'sticky mittens' is placed on the infant's hands. These mittens have velcro attached, which allows the infant to pick up and interact with block objects that have also been covered in velcro. The child can learn to pick up objects simply by swiping their arms around. The infants in the study are three months olds; at this age, they have not yet learned to grip things precisely. The crucial finding is that, when the mittens are subsequently removed, the infants who have had experience with the mittens spend more time exploring and manipulating objects relative to a control group who have not experienced the mittens: experience with the mittens appears to advance the infant's 'knowing' of objects (Needham et al., 2002). This study is itself evidence that infants learn about object affordances, or first-order relations, through active exploration of the environment.

8.2.2 Design

An area where the relational field model might be put to more immediate practical use is in the design of objects and places. It may be possible to generate a useful set of conceptual tools by thinking of design in terms of relations between objects and users of those objects.

Donald Norman, in his book *The Psychology of Everyday Things* (1988), talks of a design principle which he calls ‘natural mapping’. This refers to a relationship between a piece of technology and the set of controls for operating it: gas knobs on a stove are easier to use if they are laid out in the same configuration as the burners that they control, rather than arranged in a straight line (see Fig. 8.1); similarly, light switches are easier to use if the layout of the switches corresponds to the layout of the actual light fittings in the room. Norman suggests that this design principle works because in using natural mapping, designers are ‘taking advantage of physical analogies and cultural standards’ (p. 23). It is unclear exactly what Norman wants to suggest about the nature of this analogy: is a ‘physical analogy’ a directly perceivable relation between two things in the world, or is it a relation between the mental representations of those two things? Norman may have intended the latter, but the description is consistent with the former, and is therefore amenable to an explanation in terms of the relational field model. On the field model, there exists a set of perceivable relations between the controls and the thing being controlled. Gas knobs that are laid out in the same configuration as the burners are related to those burners by a geometric transformation (an affine transformation). It is unsurprising that this makes the meaning of the controls easy to apprehend: such transformations are the kind of thing that our visual system deals with all the time, as we walk around objects (Michaels and Carello, 1981). The straight line of knobs, by contrast, is an impoverished layout because it destroys (or fails to make visible) the structure that specifies how the controls are transformed to fit the burners.

Another of Norman’s design principles instructs designers to ‘make things visible’. This can now be refined into a more specific directive: make relations visible. Particularly, make visible the relations that are relevant to a perceiver-actor seeking to get something done. And this applies not only to designing objects like light switches,

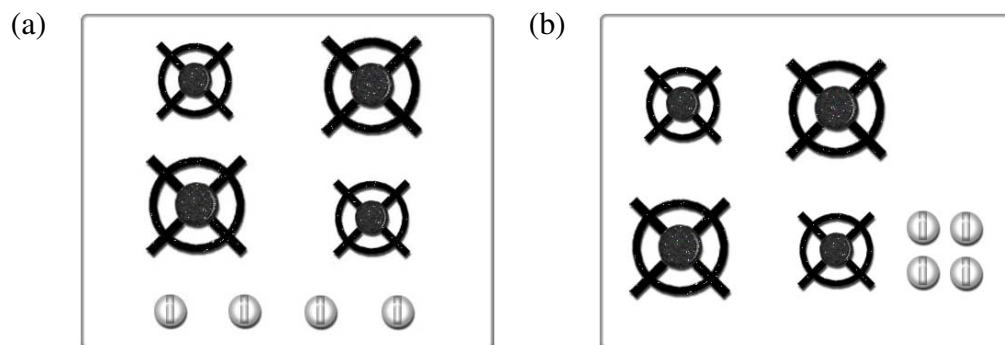


Figure 8.1: If gas knobs are arranged in a line as in (a), it is not clear which knob operates which burner; in (b), the layout of the knobs is related to the layout of the burners through a simple geometric transformation—the relations are made visible through the use of ‘natural mapping’ (images from Wikimedia Commons)

but also to designing spaces where multiple individuals interact, such as public spaces in cities. Relational design principles would emphasize how individuals can be given perceptual access to the structure that they need to guide their movement around a built environment.

One potential domain of application is in urban design. Consider one school of thought that has arisen among urban designers: the ‘shared space’ approach ([Hamilton-Baillie, 2008](#)). This says that traffic signs and pavement clutter are generally to be discouraged on the grounds that they prevent informative direct interaction between road users. Among the things to be avoided are pedestrian crossings that display red and green ‘walk’ signals. A pedestrian looking at a green ‘walk’ signal cannot simultaneously monitor the road itself for actual dangers. Like the gas knobs arranged in a line, the green man signal is an impoverished structure. It distracts attention from the rich, meaningful set of transformations that specify objects moving relative to the pedestrian (objects which, significantly, afford injury to the pedestrian) and directs attention instead to a binary instruction: ‘go’/‘don’t go’. And the same applies to drivers: the need to attend to signals apparently undermines the driver’s responsibility to monitor the actual environment. Removing these signals forces direct interaction between road

users, perhaps through eye-contact or other gestures.

This way of thinking about urban design has become popular in the United Kingdom in recent years. There are, however, some serious issues with many of resulting road schemes (Moody and Melia, 2014). An obvious problem with shared space is that it relies on visual perception and therefore potentially renders the space unusable by blind people (Thomas, 2011). Blind people rely on a physical separation between pedestrians and traffic for their safety, and they make use of kerb edges as a source of proprioceptive information about where they are in space (they make use of haptic structure in the same way that sighted people make use of visual structure that specifies the pavement edge). Haptic exploration of the kerb edge is instrumental in how blind people use the space. If the shared space design philosophy advocates the removal of kerb edges (because kerbs are deemed to be ‘clutter’), it must replace this structure with an alternative resource that is equally informative for blind users of the space. This may not be an insurmountable problem for designers. A possible solution might be to use drainage channels set into the ground in place of kerbs raised from the road (Childs et al., 2010). These could serve to inform the blind pedestrians of where the road edge is. However, some other innovation would still be necessary to allow blind people to cross in the absence of light signals.

A fundamental objection to the concept of ‘shared space’ is that it makes no allowance for the fact that different groups of road users are fundamentally unequal. There is a hierarchy of menace among road users. A moving car or truck affords serious injury to a cyclist or pedestrian, but the opposite is not true: a fast-moving pedestrian affords no particular threat to the truck-driver’s safety. This inequality means that, for road with high traffic volume, the space is never likely to be shared, and such schemes are unlikely to encourage the uptake of cycling, as advocates of shared spaces claim they will (Hamilton-Baillie, 2008).

Urban design is not normally described in terms of relations between actors and environmental structure. The concept of shared space, indeed, is largely based on designers’ intuitions about what works. What is lacking is a firm grounding in a coherent psychology of perception.

The relational field model may be able to provide the necessary framework. Consider again the task of crossing a road. Whether it is safe to cross a road becomes

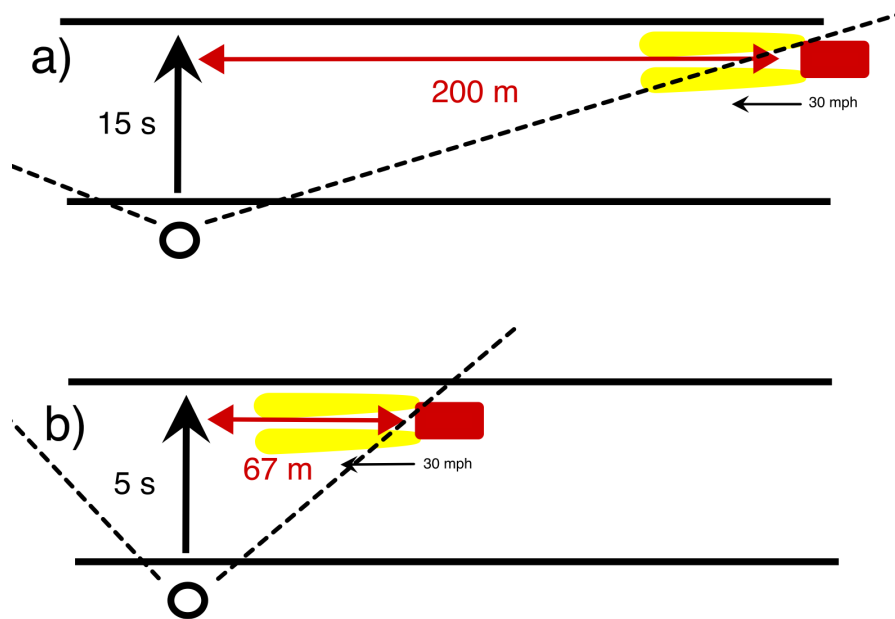


Figure 8.2: The black circles here represent pedestrians attempting to cross the road. The difficulty of the perceptual decision task increases in proportion to the time it is *going to take* for the pedestrian to cross the road. In (a), a slow-moving pedestrian, for whom the crossing will take 15 seconds, must be attentive to the possibility of vehicles approaching from 200 metres down the road. In (b), a faster pedestrian crossing the same road, but able to do so in one third of the time, need only be attentive to vehicles within a much narrower cone.

increasingly harder to judge as the distance to be crossed increases (see Fig. 8.2). One thing that makes things difficult is that it takes time to cross, and the pedestrian needs to be able to see whether there are any cars moving on the road that are on a path such that the pedestrian's path and the car's path will collide at some point. This judgement need not involve internal calculation about future events. There is information that specifies how much time the pedestrian has to cross in front of vehicles that are moving and visible. The problem is that it will not necessarily be the case that all potentially haz-

ardous vehicles are visible to the pedestrian from their starting point. A target vehicle may be presently occluded behind other objects, because of traffic, or because the road bends, say. The further the end point is from the start point, the more uncertainty there is. This uncertainty will also be specified in the optic array: a pedestrian, in learning to cross roads, must learn to attend to relations between points along the prospective path and vehicles on a course to cross through those points. A straightforward property of these relations is that vehicles travelling on the far side relative to the pedestrian can travel from further away and intercept the pedestrian's path, compared to vehicles on the near side travelling at the same speed. In effect, there exists a cone of potential hazards from the pedestrian's point of view: there is information that specifies possible dangers relative to the prospective path.

Some straightforward design principles follow. At any point in a road at which a pedestrian may wish to cross, the distance to be crossed should be made short enough such that the pedestrian's cone of potential hazards extends only into visible space. The cone should not encompass possible vehicles in locations that, from the starting point, are occluded behind other permanent surfaces. In general, crossing distances should be made as short as possible, through the use of traffic islands or central dividers.

These seem like common sense observations, and they are. The point is that the ecological framework is able to specify this common sense with sufficient precision such that it can be applied consistently in real situations. For instance, the cone of potential hazards could quite easily be implemented in computer simulations to help design actual road layouts. A computer simulation would make it possible to check, for an arbitrary number of possible starting points along the length of the road, whether sufficient information is available to the pedestrian to cross safely.

Road crossing is just one task that people engage in in a populated space. The way that these spaces are designed impacts on the degree to which individuals are able to carry out the tasks they wish to carry out. Thinking about these tasks in terms of relations has the potential to provide clear and practical design insights.

8.3 Exploring the field

The main aim of the thesis has been to develop a coherent programme for studying social phenomena—phenomena that involve coordination between multiple individuals—on ecological realist principles. The relational field perspective offers a way of understanding these interactive phenomena while avoiding some of the standard pitfalls: group minds, inward-facing regresses, solipsistic doubt about the intentions of others, and the anthropomorphic tendency to assume that social animals must really be achieving their behaviour through the use of an internal language. But, as I have stressed repeatedly, these theoretical claims will count for nothing if they cannot generate a progressive programme of empirical research. Pursuing this programme will require sustained and meticulous engagement with how animals make use of their environment and of each other. I believe this field merits further exploration.

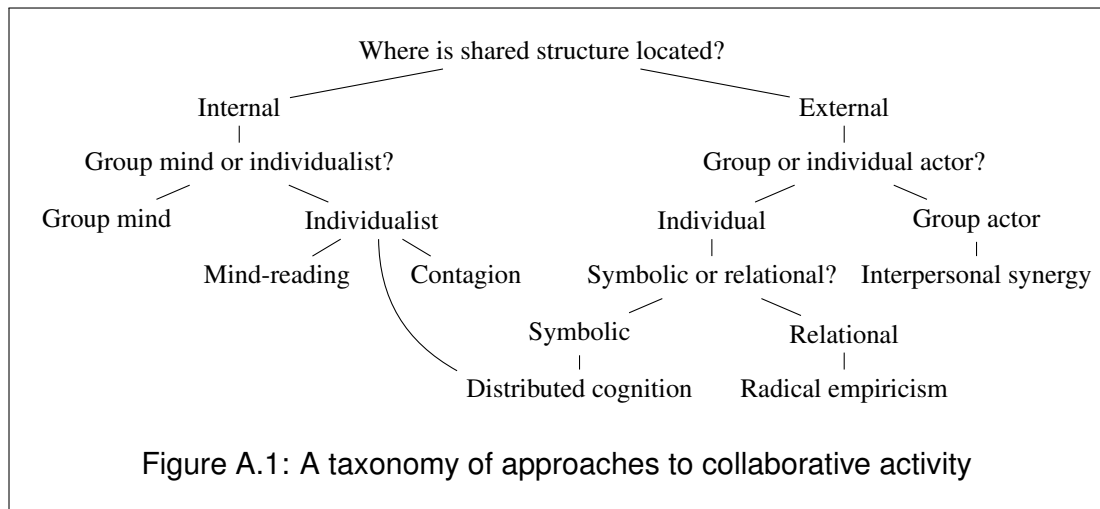
Appendix A

A taxonomy of approaches to the analysis of collaborative activities

The purpose of this appendix is twofold. First, it places the radical empiricist approach to collaborative activities (as developed in the text) in contrast with existing approaches, making explicit some of the key differences at the level of basic assumptions. Second, it identifies, very briefly, some of the major limitations and obstacles faced by the existing approaches. The point here is not to dismiss those existing approaches out of hand, but merely to suggest that the existing approaches might come up against problems that the radical empiricist approach is able to side-step. Of course, to give a comprehensive treatment of all of the rival approaches would take up many chapters. I have kept the discussion here very brief. I have found it necessary to place this material here in an appendix in order to avoid the text itself from becoming overly disputatious and preoccupied with small differences.

A.1 Overview

Collaborative activities pose a conceptual problem for psychologists. Standard psychological methodology privileges the individual as the unit of analysis. But a collaborative activity is one that involves more than one individual. If psychologists wish to talk about collaborative behaviour we need to have clear answers to two questions: 1) where is shared structure located in the system: internal to individuals' minds, or out in



the world, or both? and 2) should the group be considered an actor in its own right, or should we stipulate that only individuals can act? Responses to these questions define a taxonomy of five logically distinct approaches to the study of collaborative activities: standard cognitivism, group mind theories, distributed cognition, interpersonal synergy or group actor theories, and radical empiricism. It is important to be aware of the differences between these approaches if we are to avoid confusion when discussing group behaviour.

Psychologists have long struggled with the question of how to understand the relationship between the individual actor and the group. An individual person with an individual brain can play chess, but how does an individual soccer team coordinate its actions given that it possesses multiple brains and everyone is running around? It is tempting to think that there are basically two extreme positions that one could take on this, and that these constitute two ends of a continuum. Solomon Asch's 1952 textbook on *Social Psychology* makes exactly this argument; he contrasts two theses: 'The individualistic thesis ... begins with the assertion that individuals are the sole reality and often ends with the denial of the reality of groups. ... In extreme opposition to the preceding view is the group mind thesis. It begins with the observation that when men live and act in groups there arise forces and phenomena that follow laws of their own and which cannot be described in terms of the properties of individuals composing them.'

But this oversimplifies things. For one thing it leaves out any mention of the structure of the shared world in which group activities take place. In reality there is not merely a single continuum. Fig. A.1 presents a more detailed map of the logical space. Any attempt to incorporate collaboration into a psychological framework must give an account of how the system made up of environment and collaborating actors is assembled. And further, for collaborative activity to be possible, it is necessary that some structure in this system, in the form of either mental representations or perceptual sensitivity to physical structure, must be shared between actors. Two main questions follow: 1) where is this shared structure located in the system: internal to individuals' minds, or in the environment, or both? and 2) should the group be considered an actor in its own right, or should we stipulate that only individuals can act?

The branches on the left hand side in Fig. A.1 correspond to the set of traditional cognitivist approaches to collaborative activities (this is Asch's continuum). These are characterized by the assumption that actions are the products of minds, be it the mind of an individual or of a group. On the right hand side are the non-representational approaches, which conceive of actions as consequences of the dynamics of actor-environment systems. The distributed cognition approach emerges here as a hybrid, combining elements from both branches.

A.2 Internalist approaches

A.2.1 The group mind

What we might call the strong version of the group mind thesis would hold that groups are conscious entities with their own subjective experience. Though this is logically possible, one would be hard pressed to build a psychological research programme on such an idea. The weaker version states merely that it may be useful, for analytical purposes, to treat groups *as if* they possessed agency in their own right. Thus, [Theiner et al. \(2010\)](#) propose that 'specific cognitive capacities that are commonly ascribed to individuals are also aptly ascribed at the level of groups.' As an example they cite the unplanned formation of trails, or desire lines, over grass-covered areas within built environments. These trails appear as the grass gets worn down by people taking the

shortest path between buildings. The authors argue that this trail-formation cannot be ascribed to any individual, since there is no one person that deliberately sets out to wear down the grass. Are we then justified in ascribing the behaviour to the group? To say that the group creates the trails is a perfectly valid description in everyday language, but it has no clear explanatory value for psychologists' purposes. It does not allow us to predict the group's future behaviour, for instance. This glide from description to explanation was also diagnosed by Floyd Allport (1924) nearly a century ago. Allport rejected the group mind thesis with some disdain: 'The crowd mind theory is not only false; it retards in a special manner the discovery of the truth. Pointing towards the whole rather than the parts, it withdraws attention from the latter and incites thought in precisely the wrong direction.'

A.2.2 Individualism

The individualist approach seeks to locate the explanation of collaborative activities exclusively in the minds of the individual participants. This has been by far the most common type of approach in the history of psychology as a modern discipline (for a historical overview, see Tindale et al., 2001). Advocates of the individualist approach, eschewing the idea that a group can have its own autonomy, are faced with the problem of how people are able to act together at all, given that, according to the principles of this approach, individual minds are private, sealed entities that have no direct access to the contents of other minds, and have access to the world only indirectly, via impoverished sense data in the form of light beams striking a retina or sound waves tickling an eardrum, and so on. How can a group of minds, given these conditions, ever so coordinate their activities as to be able to act in concert?

There are ostensibly two types of solution that individualists propose as mechanisms for coordinating between distinct minds: mind-reading and contagion. In reality, however, there is less distance between the two positions than there appears to be at first glance.

Mind-reading

The first possible solution to the problem of other minds is to argue that individuals keep an internal record of what they believe others are thinking. This solution is called mind-reading, or mentalizing, theory of mind, meta-representation, or other-modelling. At present this is often taken as the standard mainstream explanation against which all other approaches to social cognition should be measured (e.g., [Overgaard and Michael, 2013](#)). Given this privileged position, one might expect that the mind-reading view must coalesce around some core agreed set of assumptions. In fact, the opposite is the case. The one thing that is widely agreed upon is that there exist two competing accounts of what mind-reading is: the theory theory and the simulation theory ([Gallese and Goldman, 1998](#)).

According to the theory theory, we are able to act collaboratively because we have explicit hypotheses about the contents of each other's minds. Everyday understanding of others' behaviour is thus held to be achieved in much the same way that an experimental psychologist goes about trying to understand the behaviour of a research subject ([Gopnik and Wellman, 1992](#)). The competing simulation theory says that instead of formulating explicit internal hypotheses, the mind builds an internal model or simulation of the person being observed ([Goldman, 1992](#)). This account has the attractive quality of making the understanding of others' actions continuous with the understanding of one's own actions: if mind-reading is internally simulating someone else's behaviour, then understanding a given behaviour may perhaps recruit the same internal machinery as is used when actually carrying out that behaviour. To understand others is to put oneself in their place—to imagine, possibly subconsciously, carrying out the observed action oneself.

Contagion

Mind-reading is presented as a mechanism by which mental content is shared. The contagion theory makes the more eye-catching claim that it is not only mental content that is transmitted, but behaviour as well, and that this happens without our awareness. Thus, it is claimed that we are more likely to drive too fast if we've been watching motor racing ([Dijksterhuis and Bargh, 2001](#)), that we unconsciously repeat each other's

words and syntactic constructions (Pickering and Garrod, 2004), and that obesity is transmitted through social networks like a disease (Christakis and Fowler, 2007).

The notion of contagion here is ambiguous, however. It is necessary to distinguish between two types of contagion theory: a naïve behaviour-transmission model, and a more sophisticated content-transmission model. The behaviour transmission model predicts that merely being exposed to a behaviour means that an individual will begin copying that behaviour. The content-transmission model instead states that while mere exposure may lead automatically to *internal* copying or covert simulation, the *outward* behaviour of the perceiving individual is not determined by this simulation; the stimulation instead acts as one influence on behaviour among others.

Do any of these authors genuinely believe in the first of these, the naïve behaviour-transmission model? It turns out, when one reads them carefully, that the answer is no. Any statements to the effect that perception directly determines behaviour is invariably hedged elsewhere in the text, with the result that the proposal has to be read as a content-transmission model after all.¹ The content-transmission model is not clearly distinct from the simulation theory of mind.

The naïve behaviour-transmission model does in theory make clear predictions about individual behaviour. The content-transmission model makes only probabilistic claims about populations. This is one limitation of the approach. It cannot reliably predict behaviour in a specific individual: if a given individual fails to copy the overeating behaviour of his or her peers, we cannot tell whether that is evidence against the contagion model itself, or merely evidence that this particular individual is not easily susceptible to such influence.

¹One example, from Dijksterhuis and Bargh (2001): early on, the authors state that imitating others ‘flows directly from a fact of mental representation and organization—that perceptual and behavioral representations for the same action overlap. [...] It is not necessary that the behavioral response be stamped in as a habit through reinforcement and it is not necessary for the response to be intended and strategic.’ But later on: ‘However, we also argued that these effects could sometimes be inhibited or moderated. Without the possibility to moderate direct effects of perception on behavior, we would indeed behave like the fish or frogs discussed earlier.’

A.3 Distibuted cognition

Distributed cognition, under the scheme being put forward here, is a hybrid approach which is internalist about some things and externalist about others. It inherits cognitivist assumptions about the computational nature of mind, while acknowledging the importance of the external structure that underlies the performance of tasks as we actually carry them out.

I will focus on one concept which is central to the distributed cognition approach and which distinguishes it from the other perspectives under discussion. This is the concept of *external representations*. The idea is that, in addition to representing the world internally, in the form of mental symbols or internal representations, we also represent the world *externally*, e.g. in the form of diagrams, maps, written instructions; rather than storing all of this information internally, we store it in the environment and consult it when necessary. The suggestion is that these external artifacts are ubiquitous in the activities we routinely engage in.

This concept of external representations is a major strength of the distributed cognition perspective. It makes it possible for analysts to construct compelling descriptions of team behaviour in activities such as ship navigation, which involves various maps and instruments (Hutchins, 1995). One of the stated functions of external representations is that they can serve as ‘a sharable object of thought’ (Kirsh, 2010). Two people looking at a map can refer to things on the map directly by pointing at them, presumably without having to take the other’s perspective into account in any deliberate manner.

There is an ambiguity here: while ‘external representation’ is a useful label to give to a map, it can be unclear what place the actual *territory* has within the scheme. Is the territory a mere source of input for the representations—is it the thing that the representations have to correspond to? Or might we consider the territory as a kind of representation in its own right?

The latter is occasionally implied. Hutchins, in discussing Micronesian navigation and comparing navigation by stars to navigation by sighting of the destination island, says that seeing an island ‘give[s] the navigator a more direct representation of where he is’ (p. 88), and that it provides a ‘constraint’ on where the boat might be relative

to the island (p. 92). More explicitly, (Kirsh, 2010), talking about music, states that: ‘Prima facie, the best representation to make sense of musical structure is music itself’. These passages seem to imply that objects can constitute representations of themselves: seeing an island as such is not sufficient; rather, in order to make sense of it we have to see the island as a *representation* of our current position relative to it. Stated like this, the representational language appears to be quite redundant here: it is not doing any work that can’t be attributed more straightforwardly to perceptual processes. A potential way out of this confusion is to define the concept of external representations in a more narrow way than has been implied. We might adopt the following definition: an external representation cannot refer to just any piece of external structure, but only to artificial structure that is brought into being in order to carry information about some other piece of structure elsewhere.

A.4 Externalist approaches

Fully externalist perspectives reject the concept of mental content outright. Here, interpersonal coordination is held to be achieved not through any kind of mental gymnastics but through perceptual processes and through the inherent dynamics of animal-environment systems.

A.4.1 Interpersonal synergies

One type of externalist perspective places an emphasis on the group as an acting system, a system that is said to be held together by dynamical coordination processes. The group is described as an ‘interpersonal synergy’ (Riley et al., 2011). A synergy is a concept derived from the study of motor control in the individual. An individual organism is made up of a set of muscles and joints that provide the motor system with a large set of degrees of freedom; this is said to give rise to a ‘degrees of freedom problem’: in order to control action in a given task, the organism must somehow reduce the freedom inherent in its musculo-skeletal system by selecting a particular action (Bernstein, 1996; Latash, 1996). The standard example is that in wielding a hammer, an individual must keep all of the arm muscles tense except for the one that is implicated

in the necessary vertical motion: the degrees of freedom of the arm have to be reduced to one. The one-degree-of-freedom system that is thus assembled is referred to as a ‘synergy’. The concept of an interpersonal synergy, then, is an attempt to extend this description so that it can be applied to groups.

The idea here is that, in order for a set of organisms to behave as a group, they must in fact *be* a group, in the sense that they must constitute a coherent, if temporary, system, with boundaries that may be ‘fuzzy’ and difficult to identify, but that are nevertheless real (Marsh et al., 2009). This interpersonal synergy perspective we might equally refer to as the group actor perspective. Like the group mind approach, it says that the whole is more than the sum of its parts, the parts in question being limited to the individual actors, and the whole in this case being an acting system that emerges at a level of organization above the individual. Silberstein and Chemero (2012) talk about 1/f noise as a measurable signature of a complexly assembled system. 1/f noise is a feature of systems with interaction-dominant dynamics, which means that the overall activity of the system is attributable more to the assembly of the system than to the assembly of the component parts. It is argued that a multi-actor system can exhibit interaction-dominant dynamics just as much as a single organism can.

This raises a specific research question: what are the mechanisms that unite individual actors together to create group actors, or to create group-level systems that exhibit interaction-dominant dynamics? In pursuit of this question, researchers in this tradition have focused on automatic entrainment phenomena, similar but at a lower level to those discussed in the Contagion section above—things like spontaneous interpersonal synchronization in the lighting patterns of fireflies and in crowds of people clapping (Marsh et al., 2006).

A.4.2 Radical empiricism

Finally, we come to the radical empiricist perspective, which, like the group actor perspective, is externalist about structure; unlike that perspective, though, it does not postulate the existence of an emergent group-level actor as a causal factor in the carrying out of collaborative tasks. Instead, radical empiricism postulates the reality of *relations* between external entities, and posits that these relations are accessible to the

perceptual systems of individual organisms (Charles, 2011; Heft, 2014; Holt, 1915). Relations are real, it says. And if this is true, then it may be possible that it is this that underpins the ability of an individual organism to engage in tasks with others who appear in its environment. If an individual is able to directly *perceive* that one of its fellows intends to carry out some action (where an intention is understood to be a relation that stands between the fellow organism and some other piece of external structure), then this may be all that is needed for the individual to be able to jointly engage in the task at hand. Notice that this obviates the need for the transmission of any internal structure between the organisms, or for any elaborate mechanism whereby organism A mentally ‘puts itself in the position of’ organism B (as the mind-reading approach has it), or for the existence of any kind of emergent group mind or group actor with an autonomy of its own, above and beyond that of its component parts.

The question that arises instead is over what exactly is the nature of these relations that are said to be directly accessible to perceptual processes. This is a question that does not admit of a simple answer. Rather, the question motivates an empirical project in which the attempt to discover the mechanisms involved in specific activities simply *is* the search for the relations underpinning those activities (see chapter 4). This project requires a methodological commitment to seeking the necessary structure only in the richly structure relational patterns, or ecological information, that connect an animal to its environment, where the environment is understood in the ecological fashion, relative to a given animal. It is also possible, within this framework, to address the problem of learning: learning to perform a new activity is conceived as adaptive change in the organism which by definition brings a change in what the environment affords that animal. Learning is not conceived as the accumulation of mental content, but as adaptive change in accordance with the doctrine of the new ball (chapter 1).

Appendix B

Glossary

Note: The definitions given here are necessarily blunt. It seemed appropriate, in giving a set of brief definitions, to pre-suppose an ecological realist standpoint here, and not labour over possible alternative definitions for a given term. Where applicable, a reference is provided to the page or chapter in the main text where a given term is introduced. I have also included here some terms not specifically discussed in the text itself (e.g., ‘belief’), which are marked with an asterisk. These, of course, deserve more thorough treatment in their own right. They are included here to give an indication of what a more complete account of the ecological realist programme would look like, above what it has been possible to deliver in the main body of the text.

affordance — an opportunity for action; specifically a relation that exists between an experiencing organism and a piece of structure in the environment scaled such that the organism is able to act upon it (chapter 1, p. 11, ff.)

alignment — see **contagion** (alignment is an awkward synonym that suffers from a classic process–product ambiguity)

assessment/management approach — an approach to the study of animal communication which holds that the function of vocal and other kinds of signalling behaviour is to ‘manage’ the activities of fellow animals, and that this management is possible because other animals are constantly ‘assessing’ their environment for threats and

opportunities; this approach is an alternative to the **sender-receiver model of communication** (p. 51)

Baldwin effect — evolutionary adaptation within a species as a result of selection pressure created by the species' own **niche construction** activities; a standard example is the spread of lactose tolerance within western human populations as an adaptation to the invention of cow-milking technology (p. 20)

belief (*) — a label we give from the outside to describe dispositions of ourselves and others; unlike an **intention**, a belief is inherently linguistic; it makes no sense to talk about pre-linguistic infants' or non-human animals' understanding of beliefs, because these individuals have no language in which to formulate beliefs (or represent beliefs to themselves); a belief can be expressed in language in the form of a speech act that sets its own truth conditions: to say X believes that Y is to say for example that under conditions Z, X will anticipate event W (among other things, perhaps: Y may imply multiple events)

blank slate — a much-derided metaphor for the infant **mind**; many cognitivists take exception to the 'blank' part, and use this as a justification for some variety of nativism; I suggest that the problem with this metaphor is not so much that it suggests that the newborn mind is *blank*, but that it is a *slate*: a slate can only serve as a surface for symbols; organisms are not slates with bodies attached, but acting, experiencing systems that learn (p. 13); for an alternative, see **doctrine of the new ball**

brain-oriented psychology — an approach to psychology that tries to explain behaviour principally in terms of activity in the brain; in cognitivism, it is further posited that the brain can be understood as being composed of **modules** (or perhaps partially so composed); chapter 4 contrasts this approach with an ecological alternative, **task-oriented psychology**

chimpanzee — a distant cousin of ours; unfairly maligned in the cognitive literature for lacking a fully human-like **theory of mind**, or at least for not yet having evidenced

having such a thing by passing a **false belief test**; in fact, it should be impossible for any animal lacking language to pass such a test, because of the linguistic nature of **beliefs** (see above); chimps also engage in collaborative activities, such as pack-hunting (p. 28)

collaborative activity — any activity that we want to understand that involves more than one actor; it is not possible to define what does or does not count as a collaborative activity in absolute terms; methodologically we can only take a commonsense activity that we are interested in and then analyse how that activity is carried out in terms of what resources are involved and how they are assembled (it may turn out that after such an analysis we do not want to call the activity collaborative after all, for example if it turns out that only one actor is in fact instrumental in completing the task; or we may decide, after further investigation, that where we thought there was a single activity we actually want to distinguish more than one task); see chapter 2

collaborative learning (*) — the study of individual learning as a process facilitated by activities carried out in a collaborative setting

collective intention — see **shared intention**

common ground — a label we give from outside to describe beliefs that can be ascribed simultaneously to multiple people, or intentions of two individuals towards the same entity at the same time (see **shared intention**); said to be achieved through the process of ‘grouding’ (p. 94, fn.)

common knowledge — a synonym for **common ground**; this label, or description of a state of affairs, is sometimes mistaken for the state of affairs itself, leading some researchers to state that common knowledge is a necessary precondition for **shared intentions**; I take these to be really just two labels with overlapping extension

communication (*) — a label we give from outside to describe a sequence of events where an organism acts on the **relational web** of another, or where an organism cre-

ates structure in the world for a second organism to act on; when structure is created without the intention that it is *for* a second organism, we are not inclined to call this communicative, but talk instead of **stigmergy**

contagion — in internalist approaches to collaboration, a mechanism by which mental content is said to be copied from one mind to another, and which is held to be more or less automatic in nature; proposed mechanisms of contagion include **mirror neurons**, linguistic interactive alignment, and ‘emotional contagion’ (p. 169, ff.)

convention — a post hoc description of the fact that the verbal actions that we use appear to be used consistently, across time and by different speakers; it is often assumed that convention is a pre-requisite for language use; a contention of chapter 3 is that the primary phenomenon of speaking is action controlled with reference to relations, and that the apparently conventional nature of speaking is really a consequence of the way that the **relational field** is structured

convergence — see **contagion**

conversation (*) — canonically, a face-to-face activity in which two or more people act on one another’s relational webs; a highly-skilled form of speaking

cultural evolution (*) — cumulative alterations to the structure of the human niche (and perhaps that of other cultural animals), including inventions of artefacts, linguistic structures, practices, and routines

demand characteristics — a set of affordances specific to the research laboratory; a description of what an experimental task looks like *from the subject’s point of view*; a given experiment lacks validity if it creates demand characteristics that are not neutral, relative to the research hypothesis, but subtly channel behaviour either towards or away from that which is ‘predicted’ by the research hypothesis; this is a problem for research in the non-task-oriented tradition (see chapter 4; also Baggs, 2013).

dialogue — a label we give from the outside to a series of verbal actions carried out by speakers on one another's **relational web** (chapter 3)

direct perception — perception not mediated by mental pictures or other internal representations; non-inferential perception (chapter 1)

distributed cognition — a cognitivist approach to describing tasks that allows that resources used in the completion of those tasks may be distributed across individual minds as well as the environment, in the form of **external representations** (p. 171, ff.)

doctrine of the new ball — a metaphor which conceives of learning as a process of change rather than a process of accretion; the analogy here is with a cricket ball whose flight behaviour changes as it ages (p. 14); this is an alternative to the doctrine of the **blank slate**

ecological optics — the system in virtue of which the layout of a structured environment is transmitted to a perceiving organism (p. 8); this is also the basis of any **collaborative activity**: participants to the activity share access to the same structured environment

ecological validity — a term mistakenly used to refer to either **external validity**, or **representative design** (p. 83, fn.)

energy array — the medium through which structure is transmitted to an organism possessing the appropriate receptors to interpret it; for visually perceiving organisms the most important of these arrays is generally light, but structure is also transmitted through sounds, and through the diffusion of substances in the air that can be picked up as smells (p. 9)

entrainment — a label used to denote coupling of rhythmic action between individuals or between an individual and some other external event; suffers from process–product ambiguity; also used as a synonym for **contagion**

event (*) — a disruption in the invariant structure of an environment; a change in the layout of surfaces occurring over a time scale that is meaningful to an observer (see also [Chemero, 2000](#))

event structure — a pattern of disruption in the layout of surfaces in the environment which can be picked up by an observer attending to the corresponding pattern in energy arrays, e.g. sound patterns produced by speakers (p. 53)

experimenter (*) — custodian of the **laboratory**, builder of the experimental niche, and potential source of experimental confounds undermining **validity**

extended conduit — a loopy system that arises if one attempts to reconcile the sender-receiver model of speech with the fact that ‘senders’ and ‘receivers’ are also actors in an environment; chapter 5 rejects the extended conduit model in favour of a model based on the **relational field**

extended mind — a metaphor that says that an individual’s mind can also encompass external objects, such as a notebook or a mobile phone; a quarrel exists in the philosophical literature between two entrenched camps, one camp insisting that this metaphor is a fallacy, that it confuses *coupling* of the mind with external objects with the idea that those external objects can *constitute* part of the mind, and the other camp accusing these people of ‘neurochauvinism’ ([Menary, 2010](#)); this quarrel does not arise on the present definition of **mind** as merely a *label* for the activities of an organism in relation to external objects: since the term is only a label, any disagreement that exists cannot be over the facts as they are in the world, but merely over what the label’s logical extension *should be*; if this is correct, then neither camp can win the coupling–constitution argument; it is suggested that within **task-oriented psychology** the argument is moot (p. 73)

external representation — in distributed cognition theory, external structure that carries information about some other piece of structure elsewhere (in information the-

oretic terms, but not in terms of the definition of **information** provided here where information has to be *for* an organism); an external representation is a **symbol**—a counterpart to a set of hypothetical internal, mental symbols (p. 171); it may be possible to talk about external structure as being representational in an ecological sense, but I try to avoid doing so as this seems likely to create confusion

external validity (*) — the degree to which we can claim that our experiment in the laboratory is representative of the real-world phenomenon we wish to model; see also **representative design**

false-belief test (*) — supposedly a test of whether an organism possesses a **theory of mind**; actually a test of whether an organism can pass a false belief test

goal (*) — a label applied from the outside that relates an animal with a future state of affairs that the animal is apparently attempting to bring about; the temptation is to assume that the animal must have some internal representation or idea about this future state, but perhaps the animal itself only needs be in a state of being directed at a particular piece of external structure

grand theory of communication — a quixotic project: an attempt to reduce the unbounded complexity of all of communicative behaviour to a simple set of rules, principles, or mechanisms; seeks to answer questions of the form What behaviour can we expect from agent A given a set of circumstances S?, instead of more tractable questions of the form What resources does agent A use to complete task T, and how are those resources assembled? The grand theory approach is an attempt to solve everything at once (p. 153)

group actor — a unit that would have to exist if **group affordances** existed; the group actor is the ecological psychologist's equivalent of the **group mind** (p. 172, ff.)

group affordance (*) — a construct sometimes proposed in an attempt to extend the concept of affordance (something that arises out of the complementary nature of an

experiencing organism and its environment, something that can be *perceived*) to the dyad or group level; implies a **group actor**

group mind — the idea that a group of individual possesses mental capacities above and beyond those of its component parts, or else that it should be treated as if it does (p. 167, ff.)

indirect perception (*) — perception mediated not merely by energy arrays but also by some other external structure or **external representation**, e.g. a diagram or a narrative about a distant event; this is ecological indirect perception, in contrast to mentalist indirect perception, where all perception is held to be indirect and mediated by mental images or mental ideas imposed on sense data; it is, however, not clear whether it will really be useful to talk of ‘indirect perception’ in ecological terms; the **relational field** model perhaps makes such talk unnecessary

information — the basis of perception; created by an organism through its own movements in a structured environment, enabling extraction invariant structure from energy arrays (p. 10)

intention (*) — a relation that exist between an organism and another entity; these relations can be directly *perceived* by others attending to the event structure created by unfolding actions (this is a rejection of the internalist conception of intentions as mental objects with some content that have to be ‘understood’ by others); intentions are entirely non-linguistic (they may in some sense be caused by, or directed towards, some piece of linguistic structure, but the term ‘intention’ here picks out only the *relation between* the organism and the structure, not the structure itself) and should not be confused with **beliefs**, which are inherently linguistic; relations can be *perceived*, only beliefs have to be ‘understood’

internal validity (*) — the degree to which we can claim that our description of what is going on in a laboratory experiment is a true description of what is actually going on from the subject’s point of view, i.e. whether we have accurately described the **de-**

mand characteristics present; absence of internal validity may be a consequence of the **psychologist's fallacy**

internal representation (*) — in mentalist theories, the symbols through which perceptual information (sense data) is transmitted to the mind, and over which mental processes operate; in the approach advocated here, such internal representations are avoided as a matter of methodological stipulation; closer in spirit to the ecological approach are **external representations**

invariant structure — the meaningful structure available in **energy arrays**; an organism can extract this invariant structure from irrelevant, **variant structure** through its movement (invariant structure causes perceived surfaces to change in a lawlike way under transformation during optic flow) and through physical exploration of that structure (such as in wielding an object to judge what action possibilities it affords); 'invariant' refers here to the structure in the array, rather than to the objects in the world (p. 5, ff.)

interpersonal synergy — see **group actor**

joint action — see **collaborative activity**

joint attention (*) — a label used from the outside to describe a state of affairs in which two or more individuals simultaneously attend to the same object or event

laboratory (*) — an artificially-constructed ecological **niche** in which an organism is temporarily known as a 'subject', and the specifics of what is afforded to this subject—the **demand characteristics** of the experiment—are carefully arranged not by nature but by a niche constructor known as the **experimenter**; failure to provide the subject with demand characteristics that are **externally valid** may be a consequence of the **psychologist's fallacy**

language — a label given to the use of a particular set of verbal actions, or to an individual speaker's capacity to carry out such actions, or to a set of such actions that

are common to a particular community; a given language might also be thought of as a **technology** that facilitates interpersonal coordination of attention; humans are apparently especially well adapted to this technology, perhaps as a consequence of the **Baldwin effect**; language (or speaking) is the topic of chapter 3

learning — change in the animal as a system that establishes new **affordance** relations with the environment (p. 155, ff.)

management — see **assessment/management approach**

meaning (*) — what a given object or situation affords for a given animal; meaning is apprehended by the animal in the course of its exploratory activities

mind (*) — a label we use from the outside to describe the intentions and activities of ourselves and others towards external objects and events

mind-reading — effortful tracking of another's perspective, often assumed to be a necessary pre-requisite for an animal's ability to engage in a collaborative activity (p. 169, ff.)

mirror neuron (*) — a neuron that fires both when an animal performs an action and when it observes another animal carrying out a similar action; mirror neurons have inspired much speculation by researchers; the hope appears to be that they provide a means of reducing collaborative activity to physical activity in an individual's brain; such speculation does run the risk of anthropomorphizing the neuron itself

module — a hypothetical sub-unit of the brain that is said to have evolved for a specific purpose, e.g. processing visual input; the module is a tool used in **brain-oriented psychology** to divide the problem space (explaining the functioning of the entire brain) into tractable units (explaining the functioning of modules); see chapter 4

new ball — see **doctrine of the new ball**

niche — the environment of a given organism—the subset of the structure of the world that is relevant to that animal’s life activities; the niche furnishes and constrains what action possibilities or affordances are available to the organism; the organism can in turn alter its niche through **niche construction**; the human niche is largely inherited through niche construction activities of previous generations

niche construction — the creation of new affordances in the environment through activity that alters the layout of that environment (p. 19)

perceptual learning — change in the animal as a system which enables it to make increasingly fine discriminations between different types of external structure (p. 13)

perturbation — when applied to an animal-environment system, perturbation occurs whenever the environment is changed (through, say, a natural disaster, or the actions of other animals), or whenever the animal itself is changed (through **learning**, or through brain injury, say), such that the pre-existing **affordance** relations is disrupted: the set of opportunities for actions has changed because the way the animal is built no longer corresponds to the way the environment is structured in the same way as before (chapter 7)

prediction (*) — an internal mechanism supposed necessary by some cognitivists for the control of future action in response to supposedly impoverished input in the present

prospective control (*) — a reason why **prediction** may not be necessary; control of present activity based on future events that are specified by present events, e.g. a driver’s control of braking is based on time until collision with an object on the highway, this time being specified by the present rate at which the gap is closing between the driver and the object

psychologist’s fallacy — mistaking our description of some behavioural phenomenon from the outside for the causal mechanism behind that behaviour; leads researchers to

overlook structure that is relevant to the completion of a task because it is assumed that the task is already well defined and we already know what subjects need to do to complete the task (p. 65)

referential communication games — a type of activity carried out in a **laboratory**, intended as a model of **dialogue**; the design of such tasks is strongly influenced by the **sender-receiver model of communication**, with subjects often being given specific roles within the activity, e.g., as either ‘sender’ or ‘receiver’ of instructions (chapter 5)

relational field — the complete, hierarchically-structured system of relations that exists in the world; the relations that make up this field can in principle be perceived and acted upon by an animal that is built appropriately (p. 45)

relational web — the particular sub-set of the relational field that is relevant to a given organism, and that surrounds that organism; to define the relational web of a given animal is to define its **niche** (p. 46)

representation (*) — in mentalist theories, the stuff of all mental activity—see **internal representations**; in **distributed cognition**, it is supposed that this mental activity can also transcend the boundaries of the skin, acting directly on **external representations**

representative design — a proposed gold standard for laboratory experiments, based on the assertion that in order to claim that a laboratory experiment has **external validity**, we must be able to show that the **demand characteristics** present in the experiment are aligned with, or representative of, the state of affairs pertaining when the task we are attempting to model is carried out in the wild (p. 83)

role — a label we use from the outside to describe what a given individual is doing in a given collaborative activity; that we are able to assign these labels to individuals may not be a guarantee that the individual has any appreciation of themselves as fulfilling that role, or indeed that the task, to begin with, has any such thing as a set of roles

among its properties (p. 90)

sender-receiver model of communication — a model based on the supposition that communicating is like packaging ideas into a series of containers which have to be unpacked at the other end; this may be a convenient model for some types of analysis (if, for instance, you are interested in the structure of the language itself, and indifferent to the question of how that language is actually used), but is a misleading basis for studying communication in general and irrelevant to the study of collaborative activities (chapter 5)

shared intention — a label we give from the outside to describe when two or more organisms simultaneously enter into a relation of intending towards a single entity; the concept of shared intention has no relevance to those inside the activity, who need only attend to others' intentions and act accordingly, but need have no appreciation that they are thereby entering into a wider 'shared intention'; the supposed existence of shared intentions is sometimes said to be diagnostic of whether an activity is genuinely collaborative (p. 23), but this appears to be based on a set of confusions resulting from the commission of a researcher's perspective fallacy (p. 150)

specification — a lawful relationship that exists between structure in the environment and structure in energy arrays such as light: the reason we are able to see the environment directly is because the structure in light is *specified* by the structure of the environment that gives rise to it (see chapter 1)

social affordance — a perhaps unnecessary term which designates an affordance provided not by an inanimate configuration of surfaces (an object), but by another animal—usually another member of one's own species; not to be confused with the notion of the **group affordance**

stigmergy (*) — a process in which external structure created through the activities of one organism is used in the activities of a second organism; this needn't be communicative, but can occur in the absence of intentions between organisms; it is what

allows ants to collectively discover food sources: an ant that has discovered a food source leaves a trail of pheromones, then other ants merely follow this trail (see also the example of trail-formation by humans, p. 167)

symbol (*) — a component in a formal system; on the relational field model, speaking, in its basic form, does not involve symbols; written language, however, does involve symbols (alphabets, words with standardized spellings, punctuation marks), and one's ability to use written language interacts, in a dialectical process, with one's speaking behaviour, so that a literate individual can also become skilled at speaking 'in full sentences'

task — a commonsense label we use to pick out a particular type of activity that can be carried out repeatedly and where completion of that activity appears to use the same informational resources assembled in the same sort of way every time; closer study of a given task may reveal hidden nuances, e.g. it may turn out that there is more than one way of assembling resources to reach the same ends (see chapter 4)

task-oriented psychology — the ecological approach to empirical investigation; the unit of analysis is the **task**: the animal-environment system is treated as a whole, while the activities within this system are artificially divided into sub-units (tasks) for the purpose of analysis (chapter 4)

theory of mind — a cognitive mechanism held by some cognitivists to be a necessary pre-condition of an individual's ability to engage in a collaborative activity; this is held on the assumption that intentions cannot be perceived but have to be created internally (either by internal theorizing about other minds or by internal simulation of others' activities); without this mechanism, it is held, individuals are unable to 'understand' the actions of others; theory of mind talk appears to be built on a failure to distinguish between non-linguistic **intentions** (which, under the definition given above, are simply relations) and inherently-linguistic **beliefs** (p. 169)

validity (*) — the degree to which we can claim that our description of a state of af-

fairs is true—that it corresponds to the brute facts of the world

variant structure — structure present in energy arrays that is not ecologically meaningful—it is not **invariant structure**; invariant structure is the target of perception, variant structure is noise; a perceiving organism has to filter the useful structure from this noise; it achieves this through exploration

vervet monkey — a species of small primate native to East Africa, which engages in vocal alarm calling behaviour, the subject of chapter 6

Bibliography

- Adolph, K. E., Cole, W. G., Komati, M., Garciaguirre, J. S., Badaly, D., Lingeman, J. M., Chan, G. L., and Sotsky, R. B. (2012). How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychological Science*, 23(11):1387–1394.
- Adolph, K. E., Eppler, M. A., and Gibson, E. J. (1993). Crawling versus walking infants' perception of affordances for locomotion over sloping surfaces. *Child Development*, 64(4):1158–1174.
- Adriaans, P. (2013). Information. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Center for the Study of Language and Information, Stanford, CA, fall 2013 edition.
- Allport, F. H. (1924). The group fallacy in relation to social science. *American Journal of Sociology*, 29(6):688–706.
- Arrow, K., Bolin, B., Costanza, R., Dasgupta, P., Folke, C., Holling, C. S., Jansson, B.-O., Levin, S., Mäler, K.-G., Perrings, C., and Pimentel, D. (1995). Economic growth, carrying capacity, and the environment. *Ecological Economics*, 15(2):91–95.
- Asch, S. (1952). *Social Psychology*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Baggs, E. (2013). Demand characteristics as a tool for evaluating the design of collaborative tasks. In Fernández, R. and Isard, A., editors, *Proceedings of the 17th Workshop on the Semantics and Pragmatics of Dialogue*, Proceedings (SemDial), pages 6–14, Amsterdam, Netherlands.
- Baggs, E. (2014). The task-oriented approach in psychology: a solution to Fodor's problem. In Bello, P., Guarini, M., McShane, M., and Scassellati, B., editors, *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pages 128–133, Austin, Texas. Cognitive Science Society.
- Baggs, E. (2015). A radical empiricist theory of speaking: Linguistic meaning without conventions. *Ecological Psychology*, 27(3):251–264.
- Baldwin, J. M. (1896). A new factor in evolution. *The American Naturalist*, 30(354):441–451.

- Bard, E. G., Hill, R. L., Foster, M. E., and Arai, M. (2014). Tuning accessibility of referring expressions in situated dialogue. *Language, Cognition and Neuroscience*.
- Barrett, L. (2011). *Beyond the Brain: How body and environment shape animal and human minds*. Princeton University Press, Princeton, NJ.
- Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. Macmillan, Oxford.
- Bartlett, F. (1958). *Thinking: An experimental and social study*. Basic Books, New York.
- Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. MIT Press, Cambridge, Massachusetts.
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2):145–204.
- Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, 6(2):183–200.
- Bernstein, N. A. (1996). On dexterity and its development. In Latash, M. L. and Turvey, M. T., editors, *Dexterity and Its Development*, pages 1–244. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Bingham, G. P., Schmidt, R., and Rosenblum, L. D. (1989). Hefting for a maximum distance throw: A smart perceptual mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3):507–528.
- Bjork, E. L. and Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In Gernsbacher, M. A., Pew, R. W., and Hough, L. M., editors, *Psychology and the real world: Essays illustrating fundamental contributions to society*, pages 56–64. Worth Publishers, New York.
- Blackledge, J. T. (2003). An introduction to relational frame theory: Basics and applications. *The Behavior Analyst Today*, 3(4):421–433.
- Bodrova, E. and Leong, D. (1996). *Tools of the Mind: The Vygotskian Approach to Early Childhood Education*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Boesch, C. (2002). Cooperative hunting roles among Tāi chimpanzees. *Human Nature*, 13(1):27–46.
- Boesch, C. and Boesch, H. (1989). Hunting behavior of wild chimpanzees in the Tāi National Park. *American Journal of Physical Anthropology*, 78(4):547–573.

- Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review*, 101(2):327–341.
- Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., and Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3):1465 – 1477.
- Brennan, S. E. and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482.
- Brennan, S. E. and Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2):274–291.
- Bril, B., Dietrich, G., Foucart, J., Fuwa, K., and Hirata, S. (2009). Tool use as a way to assess cognition: how do captive chimpanzees handle the weight of the hammer when cracking a nut? *Animal cognition*, 12(2):217–235.
- Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by nature and design*. Harvard University Press, Cambridge, Massachusetts.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press, Berkeley, California.
- Carletta, J., Hill, R. L., Nicol, C., Taylor, T., de Ruiter, J. P., and Bard, E. G. (2010). Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods*, 42(1):254–265.
- Charles, E. (2011). Ecological psychology and social psychology: It is Holt, or nothing! *Integrative Psychological and Behavioral Science*, 45(1):132–153.
- Chemero, A. (2000). What events are. *Ecological Psychology*, 12(1):37–42.
- Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, 15(2):181–195.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press, Cambridge, Massachusetts.
- Cheney, D. L. and Seyfarth, R. M. (1985a). Social and non-social knowledge in vervet monkeys. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 308(1135):187–201.
- Cheney, D. L. and Seyfarth, R. M. (1985b). Vervet monkey alarm calls: Manipulation through shared information? *Behaviour*, 94(1/2):150–166.

- Cheney, D. L. and Seyfarth, R. M. (1988). Assessment of meaning and the detection of unreliable signals by vervet monkeys. *Animal Behaviour*, 36(2):477–486.
- Cheney, D. L. and Seyfarth, R. M. (1990). *How Monkeys See the World*. University of Chicago Press, Chicago.
- Childs, C. R., Thomas, C., Sharp, S., and Tyler, N. (2010). Can shared surfaces be safely negotiated by blind and partially sighted people? In *TRANSED 2010: 12th International Conference on Mobility and Transport for Elderly and Disabled Persons*.
- Chomsky, N. (2011). Language and other cognitive systems. What is special about language? *Language Learning and Development*, 7(4):263–278.
- Christakis, N. A. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379.
- Christie, S. and Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, 11(3):356–373.
- Clark, H. H. (1996). *Using Language*. Cambridge University Press, Cambridge.
- Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. In Resnick, L. B., Levine, J., and Teasley, S. D., editors, *Perspectives on socially shared cognition*, pages 127–149. American Psychological Association, Washington, DC.
- Clark, H. H. and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1):1–39.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12):1304–1312.
- Costall, A. (1995). Socializing affordances. *Theory & Psychology*, 5(4):467–481.
- Costanza, R. (1989). What is ecological economics? *Ecological Economics*, 1(1):1–7.
- Cutting, J. E. (1982). Two ecological perspectives: Gibson vs. Shaw and Turvey. *The American Journal of Psychology*, 95(2):199–222.
- Davidson, D. (1980). *Essays on Actions and Events*. Oxford University Press, Oxford.
- Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, 3(4):357–370.
- Dhimi, M. K., Hertwig, R., and Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, 130(6):959.

- Dijksterhuis, A. and Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in experimental social psychology*, 33:1–40.
- Donaldson, M. (1978). *Children's Minds*. Fontana, London.
- Dunnette, M. D. (1966). Fads, fashions, and folderol in psychology. *American Psychologist*, 21(4):343–352.
- Ericsson, K. A. and Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, 49(8):725–747.
- Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3):363–406.
- Everett, D. L. (2012). *Language: The cultural tool*. Pantheon Books, New York.
- Fajen, B. R., Riley, M. A., and Turvey, M. T. (2009). Information, affordances, and the control of action in sport. *International Journal of Sport Psychology*, 40(1):79–107.
- Fodor, J. (2000). A science of Tuesdays. *London Review of Books*, 22(14):21–22.
- Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press, Cambridge, Massachusetts.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1):3–28.
- Fowler, C. A. and Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26(5):489–504.
- Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12):493–501.
- Garrod, S. and Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2):181–218.
- Garrod, S., Fay, N., Lee, J., Oberlander, J., and MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, 31(6):961–987.
- Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, 34(5):752–775.
- Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, 39(1):1–42.

- Gibson, E. J. and Pick, A. D. (2000). *An Ecological Approach to Perceptual Learning and Development*. Oxford University Press, Oxford.
- Gibson, J. J. (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology*, 49(3):182–194.
- Gibson, J. J. (1960). The concept of the stimulus in psychology. *American Psychologist*, 15(11):694–703.
- Gibson, J. J. (1961). Ecological optics. *Vision Research*, 1(3):253–262.
- Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton-Mifflin, Boston.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton-Mifflin, Boston.
- Gibson, J. J. (1982). Notes on affordances. In Reed, E. and Jones, R., editors, *Reasons for Realism: Selected essays of James J. Gibson*, pages 401–418. Lawrence Erlbaum, Hillsdale, New Jersey.
- Gibson, J. J. and Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, 62(1):32.
- Gierut, J. A. (1998). Treatment efficacy: Functional phonological disorders in children. *Journal of Speech, Language and Hearing Research*, 41(1):S85.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98(2):254–267.
- Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy*, 15(1):1–14.
- Goldin-Meadow, S., Seligman, M. E., and Gelman, R. (1976). Language in the two-year old. *Cognition*, 4(2):189–202.
- Goldman, A. I. (1992). In defense of the simulation theory. *Mind & Language*, 7(1-2):104–119.
- Gopnik, A. and Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7(1-2):145–171.
- Gréhaigne, J.-F., Bouthier, D., and David, B. (1997). Dynamic-system analysis of opponent relationships in collective actions in soccer. *Journal of Sports Sciences*, 15(2):137–149.
- Gréhaigne, J.-F., Godbout, P., and Bouthier, D. (2001). The teaching and learning of decision making in team sports. *Quest*, 53(1):59–76.

- Hamilton-Baillie, B. (2008). Shared space: reconciling people, places and traffic. *Built Environment*, 34(2):161–181.
- Hammond, K. R. and Stewart, T. R., editors (2001a). *The Essential Brunswik: Beginnings, explications, applications*. Oxford University Press, Oxford.
- Hammond, K. R. and Stewart, T. R. (2001b). Introduction. In Hammond, K. R. and Stewart, T. R., editors, *The Essential Brunswik: Beginnings, explications, applications*. Oxford University Press, Oxford.
- Harris, R. (1981). *The Language Myth*. Duckworth, London.
- Harris, R. (1988). *Language, Saussure, and Wittgenstein: How to play games with words*. Routledge, London.
- Haugeland, J. (1978). The nature and plausibility of cognitivism. *Behavioral and Brain Sciences*, 1(2):215–226.
- Hayes, S. C., Barnes-Holmes, D., and Roche, B., editors (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. Kluwer Academic, New York.
- Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, 7(1–3):43–64.
- Healey, P. G. T. (2008). Interactive misalignment: The role of repair in the development of group sub-languages. In Cooper, R. and Kempson, R., editors, *Language in Flux: Dialogue coordination, language variation, change and evolution*, pages 13–39. College Publications, London.
- Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Heft, H. (2013). An ecological approach to psychology. *Review of General Psychology*, 17(2):162–167.
- Heft, H. (2014). The tension between the psychological and ecological sciences: Making psychology more ecological. In Barker, G., Desjardins, E., and Pearce, T., editors, *Entangled Life: Organism and environment in the biological and social sciences*, pages 51–77. Springer, Dordrecht.
- Hendriks, P. and Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, 120(8):1887–1897.
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203(3):89–96.

- Hodges, B. (2009). Ecological pragmatics: Values, dialogical arrays, complexity, and caring. *Pragmatics & Cognition*, 17(3):628–652.
- Holt, E. B. (1912). The place of illusory experience in a realistic world. In Holt, E. B., Marvin, W. T., Montague, W. P., Perry, R. B., Pitkin, W. B., and Spalding, E. G., editors, *The New Realism: Coöperative studies in philosophy*, pages 303–373. Macmillan, New York.
- Holt, E. B. (1915). *The Freudian Wish and Its Place in Ethics*. H. Holt, New York.
- Hupet, M. and Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects? *Journal of Psycholinguistic Research*, 21(6):485–496.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, Massachusetts. MIT Press.
- Hutto, D. D. and Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. MIT Press, Cambridge, Massachusetts.
- Jablonka, E. (2002). Information: its interpretation, its inheritance, and its sharing. *Philosophy of Science*, 69(4):578–605.
- James, W. (1904). Does ‘consciousness’ exist? *The Journal of Philosophy, Psychology and Scientific Methods*, 1(18):477–491.
- James, W. (1912). *Essays in Radical Empiricism*. Longmans, Green and Co, London.
- Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind & Language*, 3(3):183–196.
- Kaschak, M. P. and Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, 43(3):508–529.
- Käufner, S. and Chemero, A. (2015). *Phenomenology: An Introduction*. Polity Press, Cambridge.
- Kendon, A. (2010). Spacing and orientation in co-present interaction. In Esposito, A., Campbell, N., Vogel, C., Hussain, A., and Nijholt, A., editors, *Development of Multimodal Interfaces: Active Listening and Synchrony*, pages 1–15. Springer, Berlin.
- Kirsh, D. (2010). Thinking with external representations. *AI & Society*, 25(4):441–454.
- Kono, T. (2009). Social affordances and the possibility of ecological linguistics. *Integrative Psychological and Behavioral Science*, 43(4):356–373.

- Krauss, R. M. and Fussell, S. R. (1996). Social psychological models of interpersonal communication. In Higgins, E. T. and Kruglanski, A. W., editors, *Social Psychology: Handbook of Basic Principles*, pages 655–701. Guilford, New York.
- Krauss, R. M. and Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4(3):343–346.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96(4):674–689.
- Kuhn, D. (1992). Piaget's child as scientist. In Beilin, H. and Pufall, P. B., editors, *Piaget's Theory: Prospects and possibilities*, pages 185–208. Lawrence Erlbaum, Hillsdale, New Jersey.
- Lake, R. (2012). *Vygotsky on Education*. Peter Lang, New York.
- Laland, K. N., Odling-Smee, J., and Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences*, 23(1):131–146.
- Laland, K. N. and Sterelny, K. (2006). Perspective: seven reasons (not) to neglect niche construction. *Evolution*, 60(9):1751–1762.
- Land, M., Mennie, N., and Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311–1328.
- Land, M. F. and Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25):3559–3565.
- Latash, M. L. (1996). The Bernstein problem: How does the central nervous system make its choices? In Latash, M. L. and Turvey, M. T., editors, *Dexterity and Its Development*, pages 277–303. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1):1–38.
- Lewontin, R. C. (1983). The organism as the subject and object of evolution. *Scientia*, 118(1–8):65–95.
- Malafouris, L. (2013). *How Things Shape the Mind*. MIT Press, Cambridge, Massachusetts.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Freeman, New York.

- Marsh, K. L., Johnston, L., Richardson, M. J., and Schmidt, R. (2009). Toward a radically embodied, embedded social psychology. *European Journal of Social Psychology*, 39(7):1217–1225.
- Marsh, K. L., Richardson, M. J., Baron, R. M., and Schmidt, R. (2006). Contrasting approaches to perceiving and acting with others. *Ecological Psychology*, 18(1):1–38.
- McAllister, J., Potts, A., Mason, K., and Marchant, G. (1994). Word duration in monologue and dialogue speech. *Language and Speech*, 37(4):393–405.
- McArthur, L. Z. and Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, 90(3):215–238.
- McBeath, M., Shaffer, D., and Kaiser, M. (1995). How baseball outfielders determine where to run to catch fly balls. *Science*, 268(5210):569–573.
- Mehta, R. D. (2005). An overview of cricket ball swing. *Sports Engineering*, 8(4):181–192.
- Menary, R., editor (2010). *The Extended Mind*. The MIT Press, Cambridge, Massachusetts.
- Merleau-Ponty, M. (1963). *The Structure of Behavior*. Beacon Press, Boston.
- Metzing, C. and Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2):201–213.
- Michaels, C. F. and Carello, C. (1981). *Direct perception*. Prentice-Hall, Englewood Cliffs, NJ.
- Moody, S. and Melia, S. (2014). Shared space: Research, policy and problems. *Proceedings of the Institution of Civil Engineers - Transport*, 167(6):384–392.
- Muro, C., Escobedo, R., Spector, L., and Coppinger, R. (2011). Wolf-pack (*Canis lupus*) hunting strategies emerge from simple rules in computational simulations. *Behavioural Processes*, 88(3):192–197.
- Needham, A., Barrett, T., and Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using 'sticky mittens' enhances young infants' object exploration skills. *Infant Behavior and Development*, 25(3):279–295.
- Neisser, U. (1976). *Cognition and Reality: Principles and implications of cognitive psychology*. Freeman, San Francisco.

- Noë, A. (2004). *Action in perception*. MIT Press, Cambridge, Massachusetts.
- Norman, D. A. (1988). *The Psychology of Everyday Things*. Basic books, New York.
- Orne, M. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11):776–783.
- Orne, M. (1969). Demand characteristics and the concept of quasi-controls. In Rosenthal, R. and Rosnow, R. L., editors, *Artifact in Behavioral Research*, pages 143–179. Academic Press.
- Overgaard, S. and Michael, J. (2013). The interactive turn in social cognition research: A critique. *Philosophical Psychology*, ahead-of-print:1–24.
- Owen, A. M., Hampshire, A., Grahm, J. A., Stenton, R., Dajani, S., Burns, A. S., Howard, R. J., and Ballard, C. G. (2010). Putting brain training to the test. *Nature*, 465(7299):775–778.
- Owings, D. H. (1994). How monkeys feel about the world: A review of *How Monkeys See the World*. *Language & Communication*, 14(1):15–30.
- Owings, D. H. and Morton, E. S. (1998). *Animal Vocal Communication: A new approach*. Cambridge University Press, Cambridge.
- Owings, D. H. and Zeifman, D. M. (2004). Human infant crying as an animal communication system: Insights from an assessment/management approach. In Oller, D. K. and Griebel, U., editors, *Evolution of Communication Systems: A Comparative Approach*, pages 151–170. MIT Press, Cambridge, Massachusetts.
- Oyama, S. (2000). *The Ontogeny of Information: Developmental systems and evolution*. Duke University Press, 2nd edition.
- Pan, J. S., Bingham, N., and Bingham, G. P. (2013). Embodied memory: Effective and stable perception by combining optic flow and image structure. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6):1638–1651.
- Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.
- Pinker, S. (2003). *The Blank Slate: The modern denial of human nature*. Penguin, London.
- Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, 107(Supplement 2):8993–8999.

- Pittenger, J. B. and Shaw, R. E. (1975). Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1(4):374–382.
- Platt, J. R. (1964). Strong inference. *Science*, 146(3642):347–353.
- Poehner, M. E. and Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3):233–265.
- Purver, M., Ginzburg, J., and Healey, P. (2003). On the means for clarification in dialogue. In Smith, R. and van Kuppevelt, J., editors, *Current and New Directions in Discourse and Dialogue*, volume 22 of *Text, Speech and Language Technology*, pages 235–255. Kluwer Academic Publishers.
- Quinn, R. and Carr, D. (2006). Developmentally appropriate soccer activities for elementary school children. *Journal of Physical Education, Recreation & Dance*, 77(5):13–17.
- Reddy, M. J. (1979). The conduit metaphor: A case of frame conflict in our language about language. In Ortony, A., editor, *Metaphor and Thought*, pages 284–324. Cambridge University Press, Cambridge.
- Reed, E. S. (1995). The ecological approach to language development: a radical solution to Chomsky’s and Quine’s problems. *Language & Communication*, 15(1):1–29.
- Reed, E. S. (1996a). *Encountering the World: Toward an ecological psychology*. Oxford University Press, New York.
- Reed, E. S. (1996b). *The Necessity of Experience*. Yale University Press, New Haven, Connecticut.
- Rietveld, E. and Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology*, 26(4):325–352.
- Riley, M. A., Richardson, M. J., Shockley, K., and Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, 2(38).
- Rogoff, B., Moore, L., Najafi, B., Dexter, A., Correa-Chávez, M., and Solis, J. (2006). Children’s development of cultural repertoires through participation in everyday routines and practices. In Grusec, J. E. and Hastings, P. D., editors, *Handbook of Socialization: Theory and research*, pages 490–515. Guilford Press, New York.
- Romaine, S. (1995). *Bilingualism*. Blackwell, Oxford, 2nd edition.
- Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tense of English verbs. In Rumelhart, D. E., McClelland, J. L., and the PDP Research Group,

- editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and biological models*, pages 216–271. MIT Press, Cambridge, Massachusetts.
- Russell, B. (1945). *A History of Western Philosophy*. George Allen & Unwin Ltd., London.
- Ryle, G. (1949). *The Concept of Mind*. University of Chicago Press, Chicago.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-Month-Old infants. *Science*, 274(5294):1926–1928.
- Schleidt, W. M. (1973). Tonic communication: continual effects of discrete signs in animal communication systems. *Journal of Theoretical Biology*, 42(2):359–386.
- Schober, M. (2006). Dialogue and interaction. *Encyclopedia of language and linguistics*, 2:564–571.
- Schober, M. and Clark, H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2):211–232.
- Schober, M. F. and Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In Graesser, A. C., Gernsbacher, M. A., and Goldman, S. R., editors, *Handbook of Discourse Processes*, pages 123–164. Lawrence Erlbaum Associate, Mahwah, NJ.
- Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences*, 4(3):321–354.
- Searle, J. R. (1978). Literal meaning. *Erkenntnis*, 13(1):207–224.
- Searle, J. R. (1995). *The Construction of Social Reality*. Free Press, New York.
- Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2):70–76.
- Seyfarth, R. M. and Cheney, D. L. (1986). Vocal development in vervet monkeys. *Animal Behaviour*, 34(6):1640–1658.
- Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, 210(4471):801–803.
- Shapiro, L. (2011). *Embodied Cognition*. New ideas in philosophy. Routledge, London.
- Shirouzu, H., Miyake, N., and Masukawa, H. (2002). Cognitively active externalization for situated reflection. *Cognitive Science*, 26(4):469–501.

- Silberstein, M. and Chemero, A. (2012). Complexity and extended phenomenological-cognitive systems. *Topics in cognitive science*, 4(1):35–50.
- Smith, L. B., Jones, S. S., and Landau, B. (1996). Naming in young children: a dumb attentional mechanism? *Cognition*, 60(2):143–171.
- Smith, L. B., Thelen, E., Titzer, R., and McLin, D. (1999). Knowing in the context of acting: the task dynamics of the A-not-B error. *Psychological Review*, 106(2):235–260.
- Sperber, D. and Wilson, D. (1995). *Relevance: Communication and cognition*. Blackwell, Oxford, 2nd edition.
- Sterelny, K. (2014). Constructing the cooperative niche. In Barker, G., Desjardins, E., and Pearce, T., editors, *Entangled Life: Organism and environment in the biological and social sciences*, pages 261–279. Springer, Dordrecht.
- Stoffregen, T. A. (2003). Affordances as properties of the animal-environment system. *Ecological Psychology*, 15(2):115–134.
- Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In Altmann, S. A., editor, *Social Communication Among Primates*, pages 281–324. University of Chicago Press, Chicago.
- Taleb, N. N. (2012). *Antifragile: Things that gain from disorder*. Allen Lane, London.
- Theiner, G., Allen, C., and Goldstone, R. (2010). Recognizing group cognition. *Cognitive Systems Research*, 11(4):378–395.
- Thelen, E., Schöner, G., Scheier, C., and Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, 24(1):1–34.
- Thelen, E. and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press, Cambridge, Massachusetts.
- Thibault, P. J. (2011). First-order languaging dynamics and second-order language: the distributed language view. *Ecological Psychology*, 23(3):210–245.
- Thomas, C. (2011). Briefing: Minimum effective kerb height for blind and partially sighted people. *Proceedings of the ICE-Municipal Engineer*, 164(1):11–13.
- Thompson, N. S. (1997). Communication and natural design. In Owings, D. and Beecher, M., editors, *Perspectives in ethology, Vol. 12: Communication*, pages 391–415. Plenum Press, New York.

- Thompson, N. S., Olson, C., and Dessureau, B. (1996). Babies' cries: Who's listening? Who's being fooled? *Social Research*, 63(3):763–784.
- Tindale, R. S., Meisenhelder, H. M., Dykema-Engblade, A. A., and Hogg, M. A. (2001). Shared cognition in small groups. In Hogg, M. A. and Tindale, R. S., editors, *Blackwell Handbook of Social Psychology: Group Processes*, pages 1–30. Blackwell, Oxford.
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge, Massachusetts.
- Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, 4(4):156–163.
- Tomasello, M. (2014). *A Natural History of Human Thinking*. Harvard University Press, Cambridge, Massachusetts.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H., et al. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5):675–690.
- Tonneau, F. (2004). Review of Relational frame theory: A post-Skinnerian account of human language and cognition edited by Steven C. Hayes, Dermot Barnes-Holmes & Bryan Roche. *British Journal of Psychology*, 95(2):265–268.
- Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology. *Ecological psychology*, 4(3):173–187.
- van Dijk, L., Withagen, R., and Bongers, R. M. (2015). Information without content: A gibsonian reply to enactivists' worries. *Cognition*, 134:210–214.
- Vygotsky, L. S. (1962). *Thought and Language*. MIT Press, Cambridge, Massachusetts.
- Vygotsky, L. S. (1978). *Mind in Society: The development of higher psychological processes*. Harvard University Press, Cambridge, Massachusetts.
- Warren, W. H. (1984). Perceiving affordances: visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5):683–703.
- Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, 113(2):358–389.
- Weber, S. J. and Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, 77(4):273–295.

- Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In Mullen, B. and Goethals, G. R., editors, *Theories of Group Behavior*, pages 185–208. Springer-Verlag, New York.
- Wilson, A. D. and Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, 4(58).
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636.
- Withagen, R., de Poel, H. J., Araújo, D., and Pepping, G. (2012). Affordances can invite behavior: Reconsidering the relationship between affordances and agency. *New Ideas in Psychology*, 30(2):250–258.
- Worgan, S. and Moore, R. (2010). Speech as the perception of affordances. *Ecological Psychology*, 22(4):327–343.
- Yule, G. (1997). *Referential Communication Tasks*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Zhu, Q. and Bingham, G. P. (2008). Is hefting to perceive the affordance for throwing a smart perceptual mechanism? *Journal of Experimental Psychology: Human Perception and Performance*, 34(4):929–943.